THE BACK AND FORTH ERROR COMPENSATION AND CORRECTION METHOD FOR LINEAR HYPERBOLIC SYSTEMS AND A CONSERVATIVE BFECC LIMITER

A Dissertation
Presented to
The Academic Faculty

By

Xin Wang

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Mathematics

Georgia Institute of Technology

August 2018

THE BACK AND FORTH ERROR COMPENSATION AND CORRECTION METHOD FOR LINEAR HYPERBOLIC SYSTEMS AND A CONSERVATIVE BFECC LIMITER

Approved by:

Prof. Yingjie Liu, Advisor
School of Mathematics
Georgia Institute of Technology

Prof. Molei Tao, Co-Advisor
School of Mathematics
Georgia Institute of Technology

Prof. Sung Ha Kang
School of Mathematics
Georgia Institute of Technology

Prof. Haomin Zhou
School of Mathematics
Georgia Institute of Technology

Prof. Zhiliang Xu
Department of Applied and Computational Mathematics and Statistics
University of Notre Dame

Date Approved: June 22, 2018

Man muss immer mit den einfachsten Beispielen anfangen.

David Hilbert

To my family.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor Prof. Yingjie Liu, without whom this thesis would not be possible. I have learned so much from his immerse knowledge on computational mathematics and numerical partial differential equations. His constant encouragement and guidance through my PhD study has been one of the main driving forces for us to pursue and continue the research topic of this thesis.

I would also like to thank my co-advisor Prof. Molei Tao, for many helpful discussions on the topics of this thesis and our other collaborated work, as well as constant support and encouragement for me to pursue my goal.

I am grateful to have my other committee members: Prof. Sung Ha Kang, Prof. Haomin Zhou, and Prof. Zhiliang Xu. Prof. Sung Ha Kang and Prof. Haomin Zhou have been very supportive to my PhD research and gave me many suggestions during my PhD study. Thank Prof. Zhiliang Xu for his helpful discussions and comments on the BFECC schemes for the Maxwell's equations.

I am also grateful to have learned from and worked with many other faculty and staff members in the School of Mathematics. In particular, I would like to thank Prof. Ionel Popescu, who is my mentor in the first year of my PhD study, for encouraging me to pursue my interested research directions. Thank Prof. Greg Blekherman, for guiding me on a reading course on sum of squares optimization.

I am really fortunate to have made many friends with my fellow graduate students in the School of Mathematics. I am so grateful to them for their friendship and support. Thank Mr. Haiyu Zou especially for his many helpful discussions on the topics of this thesis.

I would like to express my gratitude to the research funds during my PhD study. My research is supported in part by NSF grant DMS-1522585. Teaching assistantship from School of Mathematics have also been very helpful to me both financially and professionally.

I would like to thank my family–my parents and my sister. I dedicate this thesis to them.

TABLE OF CONTENTS

Acknowledgments v

List of Tables ix

List of Figures x

Chapter 1: Introduction 1

Chapter 2: Back-and-Forth Error Compensation and Correction Method for Linear Hyperbolic Systems 5

2.1 BFECC method for linear hyperbolic systems 6

2.2 Stability 8

2.3 Accuracy 10

2.4 Alternative view of BFECC method for hyperbolic PDE systems. 12

Chapter 3: BFECC schemes for Maxwell's equations 15

3.1 BFECC scheme based on central difference – one dimensional case 16

3.2 BFECC scheme based on central difference – two dimensional case 19

3.3 BFECC scheme based on central difference – three dimensional case 21

3.4 BFECC scheme based on Lax-Friedrichs scheme 22

3.5 BFECC scheme based on interpolation of central difference and Lax-Friedrichs schemes 24

3.6 Least square gradient approximation for non-rectangular grid 27

3.7 Point shift algorithm for grid generation 32

3.8 Divergence of magnetic field 35

3.9 Perfectly Matched Layer 36

3.9.1 One dimensional case 37

3.9.2 Two dimensional case 40

3.9.3 Implementation 45

3.10 Numerical examples 48

3.10.1 1D periodic solution 48

3.10.2 Comparison of BFECC central difference, Lax-Friedrichs, and θ-schemes 49

3.10.3 2D periodic solution 51

3.10.4 2D wave absorption by perfectly match layers 54

3.10.5 Scattering by a dielectric cylinder 58

3.10.6 Scattering by a dielectric object of complicated shape 60

Chapter 4: BFECC method for scalar conservation laws and a conservative BFECC limiter 63

4.1 Limiting by truncation 64

4.2 Conservative limiting 66

4.3 A conservative BFECC solver for the Vlasov-Poisson equation 69

4.4 BFECC for semi-Lagrangian finite volume scheme 74

4.5 BFECC for inviscid Burgers' equation 77

4.6 A characteristic difference BFECC scheme for convection-diffusion equation, viscous Burgers equation and KdV equation 82

Chapter 5: Conclusion 92

Appendix A: Stability and accuracy of BFECC schemes based on central difference 96

A.1 One dimensional case 96

A.2 Two dimensional case 98

References 107

LIST OF TABLES

3.1 Order of accuracy for BFECC + central difference scheme at $T = 0.6$ . . . 49

3.2 Error and order of accuracy for BFECC + central difference(CD), Lax-Friedrichs(LF), and θ-scheme ($\theta = 0.5$) at $T = 2.5$ 50

3.3 Order of accuracy for BFECC + least square θ-scheme at $T = 2.5$ 53

3.4 Order of accuracy for BFECC + least square θ-scheme at $T = 3.8$ 59

3.5 Grid refinement analysis for BFECC + least square θ-scheme at $T = 3.6$ . . 61

4.1 Grid refinement analysis. Numerical solutions at $T = 20$, $\Delta t/\Delta x = 2.2$. . . 68

4.2 Order of accuracy for Burgers equation, $\Delta t/\Delta x = 0.25$ 80

4.3 Order of accuracy for convection-diffusion equation, solution at $T = 0.2$. . 86

4.4 Order of accuracy for viscous Burgers' equation, solution at $T = 0.2$. . . 88

4.5 Order of accuracy for viscous Burgers' equation, solution at $T = 0.2$. . . 88

4.6 l^2 error and order of accuracy. KdV equation, solution at $T = 10^{-4}$. . . . 90

4.7 l^2 error and order of accuracy. KdV equation, solution at $T = 10^{-4}$. . . . 90

LIST OF FIGURES

3.1 Point shifted grids. 35

3.2 Comparison of numerical dissipation. 50

3.3 Comparison of numerical propagation speed of plane wave. 51

3.4 Grids: (a) Uniform rectangular; (b) (c) and (d) Non-orthogonal grids 53

3.5 Solution profile at different time: from upper left to lower right, it shows E_z surfaces at $t = 0, 0.2, 0.4, 0.6, 0.8, 1.0$. 55

3.6 Solution profile at different time: from upper left to lower right, it shows E_z surfaces at $t = 1.2, 1.4, 1.6, 1.8, 2.0, 2.2$. 56

3.7 Total energy of the electromagnetic field versus time. 57

3.8 Total energy of the electromagnetic field versus time, semi-log plot: at $t = 1.0$, total energy is reduced to 1% of the initial value and $t = 3$, it is reduced to 10^{-5} of the initial value. 57

3.9 BFECC + least square θ-scheme solution at $t = 3.8$. Left: contour plot of E_z; Right: surface plot of E_z 60

3.10 Slice of E_z with $y = 0.5$ at $t = 3.8$, compared with the analytic Mie solution. 60

3.11 Scattering by a object of complicated shape. Left: shape of the object; Right: contour plot of E_z at $t = 3.6$. 61

4.1 $|e_i^{(2)}| > |e_i^{(1)}|$ indicates overshooting/undershooting 66

4.2 BFECC with the conservative limiter. 68

4.3 Weak Landau Damping: entropy, energy, electric field and density function 71

4.4 Two stream instability: norm, entropy and energy 72

4.5 Two stream instability: electric field 73

4.6 Two stream instability: density function at T = 50 74

4.7 Shock tracking for the Burgers equation, comparison 1 81

4.8 Shock tracking for the Burgers equation, comparison 2 82

4.9 Propagation of a soliton solution for KdV equation. 89

4.10 Cosine solution evolves into solitons: from upper left to lower right, the figures shows numerical solutions at $t = 0, 0.125, 0.25, 0.375, 0.5$ and 0.625. 91

SUMMARY

In this thesis, we studied the Back and Forth Error Compensation and Correction (BFECC) method for linear hyperbolic PDE systems and nonlinear scalar conservation laws. We extend the BFECC method from scalar hyperbolic PDEs to linear hyperbolic PDE systems, and showed similar stability and accuracy improvement are still valid under modest assumptions on the systems. Motivated by this theoretical result, we propose BFECC schemes for the Maxwell's equations. On uniform orthogonal grids, the BFECC schemes are guaranteed to be second order accurate and have larger CFL numbers than that of the classical Yee scheme. On non-orthogonal and unstructured grids, we propose to use a simple least square local linear approximation scheme as the underlying scheme for the BFECC method. Numerical results showed the proposed schemes are stable and are second order accurate on non-orthogonal grids and for systems with variable coefficients. We also studied a conservative BFECC limiter that reduces spurious oscillations for numerical solutions of nonlinear scalar conservation laws. Numerical examples with the Burgers' equation and KdV equations are studied to demonstrate effectiveness of this limiter.

CHAPTER 1

INTRODUCTION

Hyperbolic partial differential equations form an important class of PDEs that often arise naturally from physics and engineering. Many physical systems are governed by certain conservation laws (for example, conservation of linear or angular momentum, conservation of energy, etc), and physicists study the system by writing down evolution equations based the conservation laws. Many of such evolution equations are hyperbolic PDEs. A few important examples are

- Advection equations;
- Wave equations;
- The Maxwell's equations;
- Inviscid Burgers' equation;
- Euler equations in fluid dynamics.

Mathematically, let $\boldsymbol{u} = (u_1, u_2, ..., u_s)$ be a unknown vector valued function of $t \in \mathbb{R}$ and $\boldsymbol{x} = (x_1, x_2, ..., x_d)^T \in \mathbb{R}^d$. Let $\boldsymbol{F}_j : \mathbb{R}^s \to \mathbb{R}^s$ be functions with continuous first order derivatives, i.e. $\boldsymbol{F}_j \in C^1(\mathbb{R}^s, \mathbb{R}^s)$ for $j = 1, 2, ..., d$. Then a first order PDE system

$$\frac{\partial \boldsymbol{u}}{\partial t} + \sum_{j=0}^{d} \frac{\partial \boldsymbol{F}_j(\boldsymbol{u})}{\partial x_j} = 0 \tag{1.1}$$

is said to be hyperbolic if the Jacobian A_j's of $\boldsymbol{F}_j$'s

$$A_j = \begin{pmatrix} \frac{\partial(\boldsymbol{F}_j)_1}{\partial u_1} & \cdots & \frac{\partial(\boldsymbol{F}_j)_1}{\partial u_s} \\ \vdots & \ddots & \vdots \\ \frac{\partial(\boldsymbol{F}_j)_s}{\partial u_1} & \cdots & \frac{\partial(\boldsymbol{F}_j)_s}{\partial u_s} \end{pmatrix}, \; j = 1, 2, ..., d,$$

satisfy the following condition: for any $v_1, v_2, ..., v_d \in \mathbb{R}$, $\sum_{j=1}^{d} v_j A_j$ is diagonalizable with real eigenvalues.

A linear hyperbolic PDE system can be written as

$$\frac{\partial \boldsymbol{u}}{\partial t} + \sum_{j=0}^{d} A_j \frac{\partial \boldsymbol{u}}{\partial x_j} + B\boldsymbol{u} = 0 \tag{1.2}$$

where $A_j, j = 1, 2, ..., d$ and B are $s \times s$ matrices that may depend on t and $\boldsymbol{x}$, and A_j's satisfy: for any $v_1, v_2, ..., v_d \in \mathbb{R}$, $\sum_{j=1}^{d} v_j A_j$ is diagonalizable with real eigenvalues [1].

Numerical solutions of hyperbolic PDE systems are of great use in applications such as computational fluid dynamics and computational eletromagnetics. Many numerical scheme have been developed for various hyperbolic PDE systems, a few examples are:

- finite difference schemes such as the upwind scheme, Lax-Friedrichs schemes, Lax-Wendroff scheme, Courant-Issacson-Rees schemes for the advection equations [2];
- finite difference schemes such as the Yee scheme and generalized Yee scheme for the Maxwell's equations [3, 4, 5, 6];
- finite volume schemes such as the Godunov scheme [7], van Leer's MUSCL scheme [8], ENO and WENO schemes [9, 10] for conservation laws;
- discontinuous Galerkin (DG) method for conservation laws [11, 12].

This thesis focuses on the Back and Forth Error Compensation and Correction (BFECC) method for linear hyperbolic PDE systems and nonlinear scalar conservation laws. Back

and Forth Error Compensation and Correction (BFECC) method is introduced in [13, 14] to obtain a higher order scheme based on a lower order scheme for advection equations. Given a scheme for advection equations, the idea of BFECC method is to improve its accuracy by estimating using forward and then backward advections and correcting its leading order error. Suppose $\mathcal{L}$ is a r-th order linear scheme for scalar linear advection equations, where r is an odd integer, the in general the BFECC scheme based on $\mathcal{L}$ is $(r+1)$-th order accurate, and is stable as long as scheme $\mathcal{L}$ has an amplification factor no more than 2, thus has a larger CFL number than $\mathcal{L}$ [13, 14]. In this thesis, we extend the BFECC method to linear hyperbolic systems, and show that similar accuracy and stability improvement can be achieved.

The BFECC method has been applied to level set interface computation and fluid simulations [13, 14, 15, 16, 17]. A two-step unconditionally stable MacCormack scheme and its generalization are developed in [18] for fluid simulations. The property that BFECC stabilizes even an unstable scheme (with its amplification factor no more than 2) is very helpful for systems because one doesn't have to compute the local characteristic information for constructing a low diffusion stable scheme. With the new extension to linear hyperbolic systems, we propose BFECC schemes for the Maxwell's equations which are second order accurate, easy to implement, and have larger CFL numbers than that of the classic Yee scheme [3].

Applying BFECC method to nonlinear conservation laws requires special attention to discontinuities. This is handled by solving a Riemann problem near discontinuities and applying slope limiter or flux limiter (as in the MUSCL scheme [8]), or specially designed nonlinear interpolation (as in the ENO and WENO schemes [9, 10]). We propose a conservative limiter that is based on the BFECC method and successfully apply it to solve scalar conservation laws.

The rest of the thesis is organized as follows:

In Chapter 2, we extend the BFECC method to linear hyperbolic systems. We established the stability and accuracy improvement theorems for the BFECC method. We also discuss the connection between the BFECC methods for hyperbolic systems and scalar hyperbolic equations.

In Chapter 3, we discuss a main application of the BFECC method for linear hyperbolic systems: BFECC schemes for the Maxwell's equations. On uniform orthogonal grids, we use central difference and Lax-Friedrichs schemes as the underlying schemes for the BFECC method. Order of accuracy and CFL numbers for the corresponding schemes are discussed. On unstructured grids, we present a first order scheme based on the least square local linear approximation and use it as the underlying scheme. The divergence of the magnetic field and the perfectly matched layer [19] implementation are also discussed. Numerical examples show that the scheme remains to be second order on non-orthogonal grids.

In Chapter 4, we report our effort on applying the BFECC method to nonlinear scalar conservation laws. We propose a conservative BFECC limiter for better treatment of discontinuities. A combination of the first order accurate Godunov scheme and the BFECC method with conservative limiter demonstrates better performance than the classical second order accurate MUSCL scheme. We also discuss application of BFECC method to the convection terms in Vlasov-Poisson system, viscous Burgers' equation and the KdV equation.

We conclude the thesis in Chapter 5, pointing out advantages/disadvantages of the BFECC schemes and a few interesting directions to pursue in future work.

CHAPTER 2

BACK-AND-FORTH ERROR COMPENSATION AND CORRECTION METHOD FOR LINEAR HYPERBOLIC SYSTEMS

In this chapter, we discuss the Back-and-Forth Error Compensation and Correction(BFECC) method for linear hyperbolic PDE systems. It is a natural extension of the BFECC method for scalar hyperbolic PDEs [13, 14]. The main application of this method would be new BFECC schemes for the Maxwell equations, which will be presented in Chapter 3.

In this chapter, we present the BFECC method for linear hyperbolic PDE systems. The method works for both constant coefficient and variable coefficient systems. We prove the stability and accuracy improvement theorems for homogeneous linear hyperbolic PDE systems with constant coefficients. For systems with variable coefficients, numerical examples in Chapter-4 suggest that the BFECC method also improves the stability and order of accuracy.

For linear hyperbolic PDE systems with constant coefficient in one spatial dimension, one can find proper linear transformations that diagonalize the coefficient matrix and decouple the system into a set of independent advection equations. The transformed unknown functions are called the Riemann invariants. Therefore, one can simply view such a system as a collection of advection equations, and the BFECC method for advection equations can be applied to the Riemann invariants. However, if the system has more than one spatial dimension, it is not always possible to diagonalize coefficient matrices simultaneously, and one must establish the stability and accuracy theorems for BFECC method in this hyperbolic system setting.

2.1 BFECC method for linear hyperbolic systems

Denote $\boldsymbol{u}(\boldsymbol{x}, t)$ the vector of unknown functions, where $\boldsymbol{x} \in \mathbb{R}^d$ and $t \in \mathbb{R}$ are the spatial and temporal variables. Consider a homogeneous linear hyperbolic PDE system with constant coefficients in the following form:

$$\partial_t \boldsymbol{u} + \sum_{i=1}^{d} A_i \partial_{x_i} \boldsymbol{u} = 0, \tag{2.1}$$

where $A_i, i = 1, 2, ..., d$ are real constant matrices, and any linear combination $\sum_{i=1}^{d} \alpha_i A_i$ is diagonalizable with real eigenvalues. When all the coefficient matrices A_i are symmetric, we say it is a symmetric linear hyperbolic system.

We solve this system numerically with a finite difference scheme. For simplicity of discussion, we assume a uniform orthogonal grid is used and discuss the scheme in the whole space. Denote the mesh sizes

$$\Delta \boldsymbol{x} = (\Delta x_1, \Delta x_2, ..., \Delta x_d),$$

and $\Delta t_n = t_{n+1} - t_n$ (we omit subscript n when Δt_n is the same for all n). Denote the numerical solution

$$\boldsymbol{U}_{\boldsymbol{j}}^n \approx \boldsymbol{u}(j_1 \Delta x_1, j_2 \Delta x_2, ..., j_d \Delta x_d, t_n),$$

where $\boldsymbol{j} = (j_1, j_2, ..., j_n)$ is the multi-index vector. Denote $\boldsymbol{U}^n = \left\{ \boldsymbol{U}_{\boldsymbol{j}}^n : \forall \boldsymbol{j} \right\}$ the collection of numerical solution at all grid points at the time t_n.

Suppose $\mathcal{L}$ is a numerical scheme for this system, i.e.

$$\boldsymbol{U}^{n+1} = \mathcal{L} \boldsymbol{U}^n.$$

We define $\mathcal{L}^*$ the backward update step from t_{n+1} to t_n by applying $\mathcal{L}$ to the time-

reversed system:

$$\partial_t \boldsymbol{u} - \sum_{i=1}^{d} A_i \partial_{x_i} \boldsymbol{u} = 0.$$

By applying the Back-and-Forth Error Compensation and Correction (BFECC) steps [13, 14], we obtain a new scheme $\mathcal{L}_{BFECC}$ which updates the solution in three steps:

1. **Solve forward**.
   $\tilde{\boldsymbol{U}}^{n+1} = \mathcal{L}\boldsymbol{U}^n$.

2. **Solve backward**.
   $\tilde{\boldsymbol{U}}^{n} = \mathcal{L}^*\tilde{\boldsymbol{U}}^{n+1}$.

3. **Solve forward with the modified solution at time** t_n.
   $\boldsymbol{U}^{n+1} = \mathcal{L}\left(\boldsymbol{U}^n + \boldsymbol{e}^{(1)}\right)$, where $\boldsymbol{e}^{(1)} = \frac{1}{2}\left(\boldsymbol{U}^n - \tilde{\boldsymbol{U}}^n\right)$.

$\boldsymbol{U}^n$ and $\tilde{\boldsymbol{U}}^n$ should have been the same if there were no numerical error. Therefore $e^{(1)}$ provides an estimate of the value lost during the forward step, which is then compensated to $\boldsymbol{U}^n$ before performing the final forward step. In general, for linear advection equations, BFECC can improve the order of accuracy by one for odd order schemes and also improve stabilities of the schemes (see [13, 14]). We establish similar results for systems of equations in the following theorems with the help of techniques in [20, 21, 14].

In the following discussion, we consider system (2.1) in $\prod_{i=1}^{d}[0,1]$ with periodic boundary conditions. And we assume the numerical scheme $\mathcal{L}$ is a linear scheme. Let $\Delta x_j = \frac{1}{N_j}$ for $j = 1, 2, ..., d$. The numerical solutions are then defined at any time on $\mathcal{D}_{\boldsymbol{N}} = \mathbb{Z}^d \cap \prod_{i=1}^{d}[0, N_j - 1]$, where $\boldsymbol{N} = (N_1, N_2, ..., N_d)$. Let $\mathcal{F}_{\boldsymbol{N}} = \mathbb{Z}^d \cap \prod_{i=1}^{d}[1 - N_j, N_j - 1]$ be the set for the dual indices of the finite Fourier series. Expand $\boldsymbol{U}^n$ as a finite Fourier series

$$\boldsymbol{U}_{\boldsymbol{j}}^n = \sum_{\boldsymbol{k} \in \mathcal{F}_{\boldsymbol{N}}} \boldsymbol{C}_{\boldsymbol{k}}^n e^{2\pi i \boldsymbol{k} \cdot \boldsymbol{x}_{\boldsymbol{j}}},$$

where $\boldsymbol{j} \in \mathcal{D}_{\boldsymbol{N}}$ and $\boldsymbol{x_j} = (j_1\Delta x_1, j_2\Delta x_2, ..., j_d\Delta x_d)$.

Since scheme $\mathcal{L}$ is a linear scheme, the coefficients of the Fourier series get updated as

$$\boldsymbol{C}_{\boldsymbol{k}}^{n+1} = Q_{\mathcal{L}}(\boldsymbol{k})\boldsymbol{C}_{\boldsymbol{k}}^{n},$$

where $Q_{\mathcal{L}}(\boldsymbol{k})$ is the Fourier symbol matrix for $\mathcal{L}$.

Remark Note scheme L is l^2 stable if the spectral radius $\rho(Q_{\mathcal{L}}(\boldsymbol{k})) < 1$ for all $\boldsymbol{k} \in \mathcal{F}_{\boldsymbol{N}}$ or $Q_{\mathcal{L}}(\boldsymbol{k})$ is diagonalizable and $\rho(Q_{\mathcal{L}}(\boldsymbol{k})) \leq 1$ for all $\boldsymbol{k} \in \mathcal{F}_{\boldsymbol{N}}$.

Denote $Q_{\mathcal{L}^*}(\boldsymbol{k})$ the Fourier symbol matrix of $\mathcal{L}^*$. Then Fourier symbol matrix Q_B for the BFECC scheme based on $\mathcal{L}$ is

$$Q_B = Q_{\mathcal{L}}\left(I + \frac{1}{2}(I - Q_{\mathcal{L}^*}Q_{\mathcal{L}})\right).$$

2.2 Stability

BFECC method improves the stability of the underlying scheme $\mathcal{L}$ for the scalar hyperbolic equation $u_t + \boldsymbol{v} \cdot \nabla u = 0$ [13, 14]. It increases the CFL number for conditionally stable schemes (for example, the upwind scheme) and making unstable schemes (for example, central difference scheme) conditionally stable. We generalize this property to BFECC method for linear hyperbolic systems with constant coefficients. The result is summarized in the following theorem.

Theorem 1. *Let $\mathcal{L}$ be a linear scheme for system 2.1. Suppose $Q_{\mathcal{L}}$ and $Q_{\mathcal{L}^*}$ satisfies the following conditions*

1. $Q_{\mathcal{L}^*}(\boldsymbol{k}) = \overline{Q_{\mathcal{L}}(\boldsymbol{k})}$ *for all* $\boldsymbol{k} \in \mathcal{F}_{\boldsymbol{N}}$*, where* $\bar{\cdot}$ *means complex conjugate, and*

2. $Q_{\mathcal{L}^*}(\boldsymbol{k})Q_{\mathcal{L}}(\boldsymbol{k}) = Q_{\mathcal{L}}(\boldsymbol{k})Q_{\mathcal{L}^*}(\boldsymbol{k})$ *for all* $\boldsymbol{k} \in \mathcal{F}_{\boldsymbol{N}}$*, and*

3. $\mathrm{Re}(\mathrm{Q}_{\mathcal{L}}(\mathrm{k}))$ *and* $\mathrm{Im}(\mathrm{Q}_{\mathcal{L}}(\mathrm{k}))$ *are diagonalizable with real eigenvalues for all all* $\boldsymbol{k} \in \mathcal{F}_{\boldsymbol{N}}$*, here* Re *is the real part and* Im *is the imaginary part.*

Then $|\rho(Q_B(\boldsymbol{k}))| \leq 1$ *for all* $\boldsymbol{k} \in \mathcal{F}_{\boldsymbol{N}}$ *if and only if* $|\rho(Q_{\mathcal{L}}(\boldsymbol{k}))| \leq 2$ *for all* $\boldsymbol{k} \in \mathcal{F}_{\boldsymbol{N}}$.

Proof. We first show $\lambda_i(Q_B) = \left(1 + \frac{1}{2}(1 - |\lambda_i(Q_{\mathcal{L}})|^2)\right) \lambda_i(Q_{\mathcal{L}})$ under the assumptions in the theorem, where $\lambda_i(Q_{\mathcal{L}})$ and $\lambda_i(Q_B)$ are eigenvalues of $Q_{\mathcal{L}}$ and Q_B, respectively.

Let $X = \mathrm{Re}(\mathrm{Q}_{\mathcal{L}})$ and $Y = \mathrm{Im}(\mathrm{Q}_{\mathcal{L}})$. Since $\bar{Q}_{\mathcal{L}} Q_{\mathcal{L}} = Q_{\mathcal{L}} \bar{Q}_{\mathcal{L}}$, we have

$$(X - iY)(X + iY) = (X + iY)(X - iY) \Rightarrow XY = YX$$

Since X and Y are diagonalizable with real eigenvalues and they commute, there is a basis set of real eigenvectors $\{v_i\}_{i=1,2,...,n}$ that diagonalizes X and Y simultaneously. Then v_i's are also eigenvectors of $Q_{\mathcal{L}}$ and $\bar{Q}_{\mathcal{L}}$, and the corresponding eigenvalues are complex conjugate of each other, i.e. $\lambda_i(\bar{Q}_{\mathcal{L}}) = \bar{\lambda}_i(Q_{\mathcal{L}})$ for $i = 1, 2, ..., n$.

By the assumption $Q_{\mathcal{L}^*} = \bar{Q}_{\mathcal{L}}$, we get $Q_B = Q_{\mathcal{L}} \left(I + \frac{1}{2}(I - \bar{Q}_{\mathcal{L}} Q_{\mathcal{L}})\right)$, and thus

$$\lambda_i(Q_B) = \left(1 + \frac{1}{2}(1 - |\lambda_i(Q_{\mathcal{L}})|^2)\right) \lambda_i(Q_{\mathcal{L}})$$

for $i = 1, 2, ..., n$.

Let $\zeta = |\lambda_i(Q_{\mathcal{L}})|$.By checking the monotonicity of function $f(\zeta) = |1 + \frac{1}{2}(1 - \zeta^2)|\zeta$ for $\zeta \in [0, \infty)$, we see $|f(\zeta)| \leq 1$ if and only if $\zeta \leq 2$, i.e. $|\lambda_i(Q_B)| \leq 1$ if and only if $|\lambda_i(Q_{\mathcal{L}})| \leq 2$, therefore the conclusion of the theorem follows. □

Remarks

1 Under the assumption of the theorem, Fourier symbol matrix Q_B has a complete (real) eigenvector basis, so $|\rho(Q_B)| \leq 1$ implies l^2 stability.

2 We comment on the assumptions of the theorem. Assumption 1 follows the same assumption in the BFECC method for advection equations [13] [14], assumption 2 requires the scheme treats backward temporal direction the same as forward temporal direction, and assumption 3 usually follows from the diagonalizability of coefficient matrix of the system. In particular, these assumptions on $Q_{\mathcal{L}}$ and $Q_{\mathcal{L}^*}$ are satisfied

for several classical schemes. Consider the following one dimensional hyberbolic system

$$\partial_t \boldsymbol{u} + A\partial_x \boldsymbol{u} = 0$$

where A is diagonalizable with real eigenvalues.

Let $\mathcal{L}$ be the central difference scheme for this system, let $\lambda = \Delta t/\Delta x$, then

$$Q_{\mathcal{L}}(k) = I - i\lambda \sin(2\pi kh)A \text{ and } Q_{\mathcal{L}^*}(k) = I + i\lambda \sin(2\pi kh)A$$

Let $\mathcal{M}$ be the Lax-Friedrichs scheme for this system, then

$$Q_{\mathcal{M}}(k) = \cos(2\pi kh)I - i\lambda \sin(2\pi kh)A,$$

and

$$Q_{\mathcal{M}^*}(k) = \cos(2\pi kh)I + i\lambda \sin(2\pi kh)A$$

It is easy to check both schemes satisfy the assumptions of the theorem.

3 A easier-to-check (but more restrictive) alternative for assumption 3 in the theorem is to require $Q_{\mathcal{L}}$ being complex symmetric. This implies X and Y are real symmetric matrices, so they are diagonalizable with real eigenvalues. We will show in Chapter-3 that this condition is satisfied for the central difference scheme and Lax-Friedrichs scheme for the Maxwell's equations.

2.3 Accuracy

We turn to the discussion for the accuracy of the BFECC method. BFECC method improves accuracy of odd order scheme for advection equations as discussed in [13, 14]. We extend this result to linear hyperbolic PDE systems with constant coefficients.

Expand the solution into Fourier series

$$\boldsymbol{u}(t,\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \boldsymbol{C}_{\boldsymbol{k}}(t) e^{2\pi i \boldsymbol{k}\cdot\boldsymbol{x}}$$

Plug into system 2.1, get

$$\begin{aligned} &\frac{\partial}{\partial t}\boldsymbol{C}_{\boldsymbol{k}}(t) = \left(-2\pi i \sum_{j=1}^{d} k_j A_j\right) \boldsymbol{C}_{\boldsymbol{k}}(t) = P(i\boldsymbol{k})\boldsymbol{C}_{\boldsymbol{k}}(t) \\ &\Rightarrow \boldsymbol{C}_{\boldsymbol{k}}(t+\Delta t) = e^{\Delta t P(i\boldsymbol{k})}\boldsymbol{C}_{\boldsymbol{k}}(t) \end{aligned}$$

where $P(i\boldsymbol{k})$ is a matrix with entries that are homogeneous linear polynomials in $i\boldsymbol{k}$ with real coefficients.

Assume $\Delta x_1 = \Delta x_2 = ... = \Delta x_d = h$, and fix $\Delta t/h$ during the mesh refinement. We first quote a theorem of Lax [22],

Theorem 2. *For a linear hyperbolic PDE system 2.1 with constant coefficients, a scheme $\mathcal{L}$ is r-th order accurate if and only if it Fourier symbol matrix $Q_{\mathcal{L}}$ satisfies*

$$Q_{\mathcal{L}}(\boldsymbol{k}) = e^{\Delta t P(i\boldsymbol{k})} + O(|\boldsymbol{k}h|^{r+1}), \text{ as } h \to 0 \text{ for all } \boldsymbol{k} \in \mathbb{Z}^d.$$

Here the $O(|\boldsymbol{k}h|^{r+1})$ term is a matrix whose entries are $O(|\boldsymbol{k}h|^{r+1})$ terms as $h \to 0$. The "only if" part is stated in theorem 2.1 of Lax's paper [22] for linear hyperbolic systems with variable coefficients. When the coefficients are constant, Lax's argument can be used to show the "if" part is also true.

We have the following theorem, which is an extension of theorem 4 in [14] to homogeneous linear hyperbolic systems with constant coefficients.

Theorem 3. *Suppose $Q_{\mathcal{L}^*}(\boldsymbol{k}) = \bar{Q}_{\mathcal{L}}(\boldsymbol{k})$ for any $\boldsymbol{k} \in \mathbb{Z}^d$ and scheme $\mathcal{L}$ is r-th order accurate for system 2.1 with constant coefficient matrices, where r is an odd integer, then the BFECC scheme $\mathcal{L}_{BFECC}$ based on $\mathcal{L}$ is $(r+1)$-th order accurate.*

Proof. Since $\mathcal{L}$ is r-th order accurate, by the Theorem-2 [22], we have

$$Q_{\mathcal{L}} = e^{\Delta t P(i\boldsymbol{k})} + Q_{r+1}(i\boldsymbol{k}h) + O(|\boldsymbol{k}h|^{r+2})$$

where $Q_{r+1}(i\boldsymbol{k}h)$ is a matrix with entries that are homogeneous degree $r+1$ polynomials in $i\boldsymbol{k}$ with real coefficients.

By the assumption,

$$Q_{\mathcal{L}^*} = \bar{Q}_{\mathcal{L}} = e^{-\Delta t P(i\boldsymbol{k})} + Q_{r+1}(i\boldsymbol{k}h) + O(|\boldsymbol{k}h|^{r+2})$$

Then

$$\bar{Q}_{\mathcal{L}} Q_{\mathcal{L}} = I + e^{-\Delta t P(i\boldsymbol{k})} Q_{r+1}(i\boldsymbol{k}h) + Q_{r+1}(i\boldsymbol{k}h) e^{\Delta t P(i\boldsymbol{k})} + O(|\boldsymbol{k}h|^{r+2})$$

Fourier symbol matrix Q_B for $\mathcal{L}_{BFECC}$ is

$$\begin{aligned}
Q_B &= Q_{\mathcal{L}} \left(I + \frac{1}{2}(I - \bar{Q}_{\mathcal{L}} Q_{\mathcal{L}}) \right) \\
&= \left(e^{\Delta t P(i\boldsymbol{k})} + Q_{r+1}(i\boldsymbol{k}h) + O(|\boldsymbol{k}h|^{r+2}) \right) \cdot \\
&\left[I - \frac{1}{2} \left(e^{-\Delta t P(i\boldsymbol{k})} Q_{r+1}(i\boldsymbol{k}h) + Q_{r+1}(i\boldsymbol{k}h) e^{\Delta t P(i\boldsymbol{k})} \right) + O(|\boldsymbol{k}h|^{r+2}) \right] \\
&= e^{\Delta t P(i\boldsymbol{k})} + \frac{1}{2} \left(Q_{r+1}(i\boldsymbol{k}h) - e^{\Delta t P(i\boldsymbol{k})} Q_{r+1}(i\boldsymbol{k}h) e^{\Delta t P(i\boldsymbol{k})} \right) + O(|\boldsymbol{k}h|^{r+2}) \\
&= e^{\Delta t P(i\boldsymbol{k})} + O(|\boldsymbol{k}h|^{r+2})
\end{aligned}$$

Therefore $\mathcal{L}_{BFECC}$ is a $(r+1)$-th order accurate scheme. □

2.4 Alternative view of BFECC method for hyperbolic PDE systems.

In some cases, we can view the BFECC method for systems as applying the BFECC method for advection equations to the Riemann invariants.

Consider an one dimensional hyperbolic PDE system with constant coefficients

$$\partial_t \boldsymbol{u} + A\partial_x \boldsymbol{u} = 0 \tag{2.2}$$

For a hyperbolic system, the coefficient matrix A is diagonalizable. Let $A = V\Lambda V^{-1}$, where Λ is a diagonal matrix with eigenvalues of A as entries, define $\boldsymbol{w} = V^{-1}\boldsymbol{u}$, then the system is equivalent to

$$\partial_t \boldsymbol{w} + \Lambda\partial_x \boldsymbol{w} = 0 \tag{2.3}$$

Variables w_i's are called the Riemann invariants of this system. Note equations for w_i's are decoupled. So we can apply the BFECC method for advection equations to each component.

Suppose now we have a r-th order scheme L for system-2.3, with r being odd, i.e.

$$\boldsymbol{W}^{n+1} = L\boldsymbol{W}^n$$

Note this scheme updates each component W_i independently from other components. Then it gives a r-th order scheme M for system-2.2,

$$\boldsymbol{U}^{n+1} = M\boldsymbol{U} := V\boldsymbol{W}^{n+1} = VL\boldsymbol{W}^n = VLV^{-1}\boldsymbol{U}$$

By theorem-4 in [14], applying BFECC to L produces a $(r+1)$-th order scheme L_B:

$$L_B = L\left(I + \frac{1}{2}(I - \bar{L}L)\right)$$

While applying BFECC to M gives us M_B

$$M_B = M\left(I + \frac{1}{2}(I - \bar{M}M)\right) = VL_BV^{-1}$$

From this expression of M_B, we see it is a $(r+1)$-th order scheme for system-2.2.

Therefore, we see in this case, the stability and accuracy improvement of the BFECC method for hyperbolic systems easily follows from the corresponding improvement for scalar advection equations, which is established in [13, 14].

Note, however, not all schemes for system-2.2 come from schemes for system-2.3 that update components of $\boldsymbol{w}$ independently, and when BFECC is applied to such a scheme, we cannot simply view it as applying BFECC to Riemann invariant independently. Also, it is numerically inefficient to decouple the system, especially when there is more than one spatial dimension. In these cases, we need theorem-1 and theorem-3 to establish the stability and accuracy improvement results.

CHAPTER 3

BFECC SCHEMES FOR MAXWELL'S EQUATIONS

Extensive studies have been done on finite difference time domain (FDTD) schemes for the Maxwell's equation [6]. Compared with other methods, for example finite element schemes, FDTD methods are very efficient, easy to understand and implement, and are able to model behaviors over all frequencies simultaneously [6]. The classical Yee scheme [3] is originally designed for uniform orthogonal grid. For non-uniform orthogonal grid, Yee scheme is known to be second order globally (though the local truncation error is first order) [23, 24]. It has been generalized to irregular nonorthogonal unstructured grids, such as the Nonorthogonal FDTD scheme [25], the Generalized Yee scheme [5] and the Overlapping Yee scheme [26]. These schemes requires and generation of nonorthogonal or unstructured staggered grids for $\boldsymbol{E}$ and $\boldsymbol{H}$ and the update rule on the unstructured grids can be complicated. In this chapter, we propose a finite difference scheme based the BFECC method that requires very few modifications when switched from uniform orthogonal grid to unstructured grids.

We first show that BFECC method turns the central difference scheme and Lax-Friedrichs scheme into stable second order accurate schemes with larger CFL number than the Yee scheme when the grid is a uniform rectangular grid. On non-orthogonal and unstructured grid, we discuss three schemes based on least square approximation.

Consider the dimensionless Maxwell's equations in a medium with zero conductivities:

$$\begin{aligned} \epsilon_r \frac{\partial \boldsymbol{E}}{\partial t} &= \nabla \times \boldsymbol{H} \\ \mu_r \frac{\partial \boldsymbol{H}}{\partial t} &= -\nabla \times \boldsymbol{E} \end{aligned} \tag{3.1}$$

Here ϵ_r and μ_r are the relative electrical and magnetic permittivity, respectively. We assume

they are constant in the following discussion.

Let $\boldsymbol{E}'(t,\boldsymbol{x}) = \sqrt{\epsilon_r}\boldsymbol{E}(\sqrt{\epsilon_r\mu_r}t,\boldsymbol{x})$, $\boldsymbol{H}'(t,\boldsymbol{x}) = \sqrt{\mu_r}\boldsymbol{H}(\sqrt{\epsilon_r\mu_r}t,\boldsymbol{x})$, then the equations for $\boldsymbol{E}'$ and $\boldsymbol{H}'$ are

$$\begin{aligned}\frac{\partial \boldsymbol{E}'}{\partial t} &= \nabla \times \boldsymbol{H}' \\ \frac{\partial \boldsymbol{H}'}{\partial t} &= -\nabla \times \boldsymbol{E}'\end{aligned}$$

To simplify discussion for the schemes, we use this simplified dimensionless Maxwell's equations in this chapter. To simplify notation, we will refer to $\boldsymbol{E}'$ and $\boldsymbol{H}'$ as $\boldsymbol{E}$ and $\boldsymbol{H}$. In this chapter, we refer to the following system as the Maxwell's equations:

$$\begin{aligned}\frac{\partial \boldsymbol{E}}{\partial t} &= \nabla \times \boldsymbol{H} \\ \frac{\partial \boldsymbol{H}}{\partial t} &= -\nabla \times \boldsymbol{E}\end{aligned} \tag{3.2}$$

Note in vacuum, we have $\epsilon_r = \mu_r = 1$, so system-(3.1) is the same as the simplified system-(3.2).

3.1 BFECC scheme based on central difference – one dimensional case

For simplicity, we consider Maxwell's equations in bounded domain $[0,1]$ with periodic boundary conditions. The dimensionless Maxwell's equations in one dimensional free space are:

$$\begin{aligned}\frac{\partial H_y}{\partial t} &= \frac{\partial E_z}{\partial x} \\ \frac{\partial E_z}{\partial t} &= \frac{\partial H_y}{\partial x}\end{aligned}$$

To simplify notation, denote $E = E_z, H = H_y$ and we have:

$$\frac{\partial H}{\partial t} = \frac{\partial E}{\partial x}$$
$$\frac{\partial E}{\partial t} = \frac{\partial H}{\partial x}$$

The central difference scheme for the above set system is:

$$\begin{aligned} \frac{E_i^{n+1} - E_i^n}{\Delta t} &= \frac{H_{i+1}^n - H_{i-1}^n}{2\Delta x} \Rightarrow E_i^{n+1} = E_i^n + \frac{\lambda}{2}(H_{i+1}^n - H_{i-1}^n) \\ \frac{H_i^{n+1} - H_i^n}{\Delta t} &= \frac{E_{i+1}^n - E_{i-1}^n}{2\Delta x} \Rightarrow H_i^{n+1} = H_i^n + \frac{\lambda}{2}(E_{i+1}^n - E_{i-1}^n) \end{aligned} \tag{3.3}$$

where $\lambda = \Delta t / \Delta x$.

With periodic boundary condition, E_j^n and H_j^n can be expanded uniquely as finite Fourier series:

$$E_j^n = \sum_{k \in \mathcal{F}_N} C_k^n e^{2\pi i k x_j}$$
$$H_j^n = \sum_{k \in \mathcal{F}_N} D_k^n e^{2\pi i k x_j}$$

where $k \in \mathcal{F}_N$ is the dual index, C_k^n and D_k^n are the Fourier coefficients for E and H, respectively.

Plug the finite Fourier series into the central difference scheme, we get:

$$\begin{pmatrix} C_k^{n+1} \\ D_k^{n+1} \end{pmatrix} = Q_{\mathcal{L}} \begin{pmatrix} C_k^n \\ D_k^n \end{pmatrix} = \begin{pmatrix} 1 & i\lambda \sin(2\pi k h) \\ i\lambda \sin(2\pi k h) & 1 \end{pmatrix} \begin{pmatrix} C_k^n \\ D_k^n \end{pmatrix}$$

where the Fourier symbol matrix $Q_{\mathcal{L}}$ is defined by the second equality. Since the spectral radius of $Q_{\mathcal{L}}$ is greater than 1, the central difference scheme is a first order scheme that is unconditionally unstable (in l^2 sense) and cannot be directly used to solve the Maxwell's equations. Applying BFECC method to the central difference scheme fixes the stability

problem and improves the order of accuracy to second order.

Solving Maxwell's equations in the backward temporal direction amount to changing λ to $-\lambda$ in the scheme, therefore we see $Q_{\mathcal{L}^*} = \overline{Q_{\mathcal{L}}}$. An easy calculation shows $Q_{\mathcal{L}^*}Q_{\mathcal{L}} = Q_{\mathcal{L}}Q_{\mathcal{L}^*}$. The real and imaginary part of $Q_{\mathcal{L}}$ are both symmetric and hence are diagonalizable with real eigenvalues. Therefore the conditions of theorem 1 and 3 are satisfied. We see the BFECC scheme based on central difference scheme is 2nd order accurate and l^2 stable if and only if $\rho(Q_{\mathcal{L}}) \leq 2$. Since the eigenvalues of $Q_{\mathcal{L}}$ are $1 \pm i\lambda \sin(2\pi kh)$, the stability condition reduces to $\max_{k\in\mathcal{F}_N}\left(1+\lambda^2\sin^2(2\pi kh)\right) \leq 4 \Leftrightarrow \lambda \leq \sqrt{3}$. Therefore the BFECC scheme based on central difference is a 2nd order accurate scheme and is stable if $\Delta t/\Delta x \leq \sqrt{3}$.

To demonstrate the details of this scheme, we did an explicit calculation for the Fourier symbol matrix, and showed it is indeed 2nd order accurate and stable if $\Delta t/\Delta x \leq \sqrt{3}$ in appendix A.

Remark In Section-3.10, we apply schemes discussed in this section to Maxwell's equations with variable permittivities. In that case, we cannot absorb the permittivities by rescaling. The schemes discussed in this section can be simply adapted to the variable permittivities case, for example, for the following system

$$\mu\frac{\partial H_y}{\partial t} = \frac{\partial E_z}{\partial x}$$
$$\epsilon\frac{\partial E_z}{\partial t} = \frac{\partial H_y}{\partial x}$$

The central difference scheme is

$$E_i^{n+1} = E_i^n + \frac{\lambda}{2\mu_i}(H_{i+1}^n - H_{i-1}^n)$$
$$H_i^{n+1} = H_i^n + \frac{\lambda}{2\epsilon_i}(E_{i+1}^n - E_{i-1}^n)$$

Here μ_i and ϵ_i are the electrical and magnetic permittivity at grid point x_i. Other schemes discussed in this chapter can be similarly adapted to the variable coefficient case.

3.2 BFECC scheme based on central difference – two dimensional case

Similar to the one dimensional case, we analyze the BFECC scheme based central difference for the dimensionless Maxwell's equation in free space in the two dimensional $\mathrm{TM_z}$ case. For simplicity, we consider computational domain $[0,1]\times[0,1]$ with periodic boundary conditions. The Maxwell's equations are:

$$\begin{aligned}
\frac{\partial H_x}{\partial t} &= -\frac{\partial E_z}{\partial y} \\
\frac{\partial H_y}{\partial t} &= \frac{\partial E_z}{\partial x} \\
\frac{\partial E_z}{\partial t} &= \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}
\end{aligned}$$

The central difference scheme is

$$\begin{aligned}
(H_x)_{i,j}^{n+1} &= (H_x)_{i,j}^{n} - \frac{\lambda_y}{2}\left[(E_z)_{i,j+1}^{n} - (E_z)_{i,j-1}^{n}\right] \\
(H_y)_{i,j}^{n+1} &= (H_y)_{i,j}^{n} + \frac{\lambda_x}{2}\left[(E_z)_{i+1,j}^{n} - (E_z)_{i-1,j}^{n}\right] \\
(E_z)_{i,j}^{n+1} &= (E_z)_{i,j}^{n} + \frac{\lambda_x}{2}\left[(H_y)_{i+1,j}^{n} - (H_y)_{i-1,j}^{n}\right] - \frac{\lambda_y}{2}\left[(H_x)_{i,j+1}^{n} - (H_x)_{i,j-1}^{n}\right]
\end{aligned}$$

where $\lambda_x = \Delta t/\Delta x$ and $\lambda_y = \Delta t/\Delta y$.

Expand H_x, H_y and E_z into Fourier series:

$$\begin{aligned}
(H_x)_{i,j}^{n} &= \sum_{(k,l)\in\mathcal{F}_{\boldsymbol{N}}} C_{k,l}^{n} e^{2\pi i(kx_i+ly_j)} \\
(H_y)_{i,j}^{n} &= \sum_{(k,l)\in\mathcal{F}_{\boldsymbol{N}}} D_{k,l}^{n} e^{2\pi i(kx_i+ly_j)} \\
(E_z)_{i,j}^{n} &= \sum_{(k,l)\in\mathcal{F}_{\boldsymbol{N}}} E_{k,l}^{n} e^{2\pi i(kx_i+ly_j)}
\end{aligned}$$

where $(k,l) \in \mathcal{F}_{\boldsymbol{N}}$ are the dual indices and $C^n_{k,l}$, $D^n_{k,l}$ and $E^n_{k,l}$ are the Fourier coefficients for H_x, H_y and E_z, respectively.

Plug into the central difference scheme $\mathcal{L}$, we get

$$\begin{aligned}
\begin{pmatrix} C^{n+1}_{k,l} \\ D^{n+1}_{k,l} \\ E^{n+1}_{k,l} \end{pmatrix} &= Q_{\mathcal{L}} \begin{pmatrix} C^{n}_{k,l} \\ D^{n}_{k,l} \\ E^{n}_{k,l} \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & -i\lambda_y \sin(2\pi l \Delta y) \\ 0 & 1 & i\lambda_x \sin(2\pi k \Delta x) \\ -i\lambda_y \sin(2\pi l \Delta y) & i\lambda_x \sin(2\pi k \Delta x) & 1 \end{pmatrix} \begin{pmatrix} C^{n}_{k,l} \\ D^{n}_{k,l} \\ E^{n}_{k,l} \end{pmatrix}
\end{aligned}$$

where

$$\begin{aligned}
Q_{\mathcal{L}} &= \begin{pmatrix} 1 & 0 & -i\lambda_y \sin(2\pi l \Delta y) \\ 0 & 1 & i\lambda_x \sin(2\pi k \Delta x) \\ -i\lambda_y \sin(2\pi l \Delta y) & i\lambda_x \sin(2\pi k \Delta x) & 1 \end{pmatrix} \\
&= I + i \begin{pmatrix} 0 & 0 & -\lambda_y \sin(2\pi l \Delta y) \\ 0 & 0 & \lambda_x \sin(2\pi k \Delta x) \\ -\lambda_y \sin(2\pi l \Delta y) & \lambda_x \sin(2\pi k \Delta x) & 0 \end{pmatrix} \\
&= I + iY
\end{aligned}$$

where $Y = \mathrm{Im}(\mathrm{Q}_{\mathcal{L}})$. Similar to the one dimensional case, solving the equation backward in time amounts to switching the signs of λ_x and λ_y in the scheme. Therefore we have $Q_{\mathcal{L}^*} = I - iY = \overline{Q_{\mathcal{L}}}$. $Q_{\mathcal{L}^*} Q_{\mathcal{L}} = Q_{\mathcal{L}} Q_{\mathcal{L}^*} = I + Y^2$. And the I and Y are both symmetric real matrices, so they are diagonalizable with real eigenvalues. We see the conditions for theorem 1 and 3 are satisfied, and hence the BFECC scheme based central difference is a 2nd order accurate scheme and is stable if $\rho(Q_{\mathcal{L}}) \leq 2$.

Compute the eigenvalues of $Q_{\mathcal{L}}$, we have

$$\lambda_1 = 1,\ \lambda_{2,3} = 1 \pm i\sqrt{\lambda_x^2(\sin(2\pi k\Delta x))^2 + \lambda_y^2(\sin(2\pi l\Delta y))^2}$$

The stability condition

$$\begin{aligned}
&\rho(Q_{\mathcal{L}}) \leq 2 \\
\Rightarrow\ &1 + \lambda_x^2(\sin(2\pi k\Delta x))^2 + \lambda_y^2(\sin(2\pi l\Delta y))^2 \leq 4 \\
\Rightarrow\ &\lambda_x^2 + \lambda_y^2 \leq 3 \\
\Rightarrow\ &\Delta t \leq \frac{\sqrt{3}}{\sqrt{(1/\Delta x)^2 + (1/\Delta y)^2}}
\end{aligned}$$

An explicit calculation of the Fourier symbol matrix for the BFECC scheme is shown in appendix A.

3.3 BFECC scheme based on central difference – three dimensional case

Results are similar as the one and two dimensional case. We can also check the assumptions of Theorem-1 and Theorem-3 in Chapter 2 are satisfied, which implies the BFECC based on central difference is second order accurate and l^2 stable if and only if

$$\Delta t \leq \frac{\sqrt{3}}{\sqrt{(1/\Delta x)^2 + (1/\Delta y)^2 + (1/\Delta z)^2}}$$

We summarize the stability and accuracy results for BFECC scheme based on central difference scheme in the following theorem

Theorem 4. *The BFECC scheme based on central difference scheme for Maxwell's equations in free space on uniform orthogonal grid is* 2*nd order accurate. It is stable in* l^2 *sense if*

1. *in one dimensional case,* $\Delta t \leq \sqrt{3}\Delta x$,

2. *in two dimensional case,* $\Delta t \leq \frac{\sqrt{3}}{\sqrt{(1/\Delta x)^2+(1/\Delta y)^2}}$, *and*

3. *in three dimensional case,* $\Delta t \leq \frac{\sqrt{3}}{\sqrt{(1/\Delta x)^2+(1/\Delta y)^2+(1/\Delta z)^2}}$.

3.4 BFECC scheme based on Lax-Friedrichs scheme

Similarly, we analyse the BFECC scheme based on Lax-Friedrichs scheme $\mathcal{M}$. In one dimension, the scheme is:

$$E_i^{n+1} = \frac{E_{i-1}^n + E_{i+1}^n}{2} + \frac{\lambda}{2}(H_{i+1}^n - H_{i-1}^n)$$
$$H_i^{n+1} = \frac{H_{i-1}^n + H_{i+1}^n}{2} + \frac{\lambda}{2}(E_{i+1}^n - E_{i-1}^n)$$

Write the one dimension Maxwell's equations as

$$\partial_t \boldsymbol{u} = A\partial_x \boldsymbol{u}$$

where

$$\boldsymbol{u} = \begin{pmatrix} E \\ H \end{pmatrix}, \; A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Then the Fourier symbol matrix the Lax-Friedrichs scheme is $Q_{\mathcal{M}} = \cos(\tilde{k}h)I + i\lambda \sin(\tilde{k}h)A$, where $\tilde{k} = 2\pi k$ is the angular wave number and $h = \Delta x$. It satisfies the conditions in Theorem-1 and Theorem-3 in Chapter 2, so the BFECC scheme based on Lax-Friedrichs scheme is second order accurate and is stable if and only if $|\rho(Q_{\mathcal{M}})| \leq 2$, i.e.

$$|\rho(Q_{\mathcal{M}})|^2 = \cos^2(\tilde{k}h) + \lambda^2 \sin^2(\tilde{k}h) \leq 4 \Leftrightarrow \lambda^2 \leq 4.$$

For two dimensional Maxwell's equations in the $\mathrm{TM_z}$ mode, write the equations as

$$\partial_t \boldsymbol{u} = A_1 \partial_x \boldsymbol{u} + A_2 \partial_y \boldsymbol{u}$$

where

$$\boldsymbol{u} = \begin{pmatrix} H_x \\ H_y \\ E_z \end{pmatrix}, \; A_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \; A_2 = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}$$

The Lax-Friedrichs scheme is

$$\begin{aligned} \boldsymbol{U}_{i,j}^{n+1} =& \frac{\boldsymbol{U}_{i-1,j}^n + \boldsymbol{U}_{i+1,j}^n + \boldsymbol{U}_{i,j-1}^n + \boldsymbol{U}_{i,j+1}^n}{4} \\ &+ \frac{\Delta t}{2\Delta x} A_1 \left(\boldsymbol{U}_{i+1,j}^n - \boldsymbol{U}_{i-1,j}^n \right) + \frac{\Delta t}{2\Delta y} A_2 \left(\boldsymbol{U}_{i,j+1}^n - \boldsymbol{U}_{i,j-1}^n \right) \end{aligned}$$

where $\boldsymbol{U}_{i,j}^n \approx (H_x(t_n, i\Delta x, j\Delta y), H_y(t_n, i\Delta x, j\Delta y), E_z(t_n, i\Delta x, j\Delta y))^T$.

Its Fourier symbol matrix is

$$Q_{\mathcal{M}} = \frac{1}{2}\left(\cos(\tilde{k}_x h_x) + \cos(\tilde{k}_y h_y) \right) I + i\lambda_x \sin(\tilde{k}_x h_x) A_1 + i\lambda_y \sin(\tilde{k}_y h_y) A_2$$

where $\tilde{k}_x = 2\pi k_x$, $\tilde{k}_y = 2\pi k_y$, $\lambda_x = \Delta t/\Delta x$, $\lambda_y = \Delta t/\Delta y$, $h_x = \Delta x$ and $h_y = \Delta y$. It is easy to check that it satisfies the conditions in Theorem-1 and 3 in Chapter 2. To compute the CFL number, we calculate

$$\begin{aligned} |\rho(Q_{\mathcal{M}})|^2 &= \frac{1}{4}\left(\cos(\tilde{k}_x h_x) + \cos(\tilde{k}_y h_y) \right)^2 + \lambda_x^2 \sin^2(\tilde{k}_x h_x) + \lambda_y^2 \sin^2(\tilde{k}_y h_y) \\ &\le \frac{1}{2}\left(\cos^2(\tilde{k}_x h_x) + \cos^2(\tilde{k}_y h_y) \right) + \lambda_x^2 \sin^2(\tilde{k}_x h_x) + \lambda_y^2 \sin^2(\tilde{k}_y h_y) \\ &\le \max\left(\frac{1}{2}, \lambda_x^2 \right) + \max\left(\frac{1}{2}, \lambda_y^2 \right) \\ &\le \max\left(1, \frac{1}{2} + \lambda_x^2, \frac{1}{2} + \lambda_y^2, \lambda_x^2 + \lambda_y^2 \right) \le 4 \end{aligned}$$

So the CFL number needs to satisfy

$$\max(\lambda_x, \lambda_y) \leq \sqrt{\frac{7}{2}} \text{ and } \lambda_x^2 + \lambda_y^2 \leq 4$$

Similarly, the CFL number satisfies the following condition for Maxwell's equations in three dimension

$$\max(\lambda_x, \lambda_y, \lambda_z) \leq \sqrt{3} \text{ and } \lambda_x^2 + \lambda_y^2 + \lambda_z^2 \leq 4$$

where $\lambda_x = \Delta t/\Delta x$, $\lambda_y = \Delta t/\Delta y$ and $\lambda_z = \Delta t/\Delta z$.

The stability and accuracy results are summerized as follows:

Theorem 5. *The BFECC scheme based on Lax-Friedrichs scheme for Maxwell's equations in free space on uniform orthogonal grid is* 2*nd order accurate. It is stable in* l^2 *sense if*

1. *in one dimensional case,* $\Delta t \leq 2\Delta x$,
2. *in two dimensional case,* $\Delta t \leq \frac{2}{\sqrt{(1/\Delta x)^2+(1/\Delta y)^2}}$ *and* $\Delta t \leq \sqrt{\frac{7}{2}}\min(\Delta x, \Delta y)$, *and*
3. *in three dimensional case,*

$$\Delta t \leq \frac{2}{\sqrt{(1/\Delta x)^2 + (1/\Delta y)^2 + (1/\Delta z)^2}} \text{ and } \Delta t \leq \sqrt{3}\min(\Delta x, \Delta y, \Delta z).$$

3.5 BFECC scheme based on interpolation of central difference and Lax-Friedrichs schemes

The Lax-Friedrichs scheme is conditionally stable but diffusive due to the average term on the right hand side, while the central difference scheme is less diffusive but unstable on itself. An interpolation between the two schemes could combine the strengths of both schemes. Let $\theta \in [0, 1]$, an θ-scheme $\mathcal{L}_\theta$ is formally $\mathcal{L}_\theta = (1-\theta)\mathcal{L} + \theta\mathcal{M}$, where $\mathcal{L}$ is the central difference scheme and $\mathcal{M}$ is the Lax-Friedrichs scheme for Maxwell's equations.

Using aforementioned notations, for one dimensional Maxwell's equation, the scheme is

$$\boldsymbol{U}_i^{n+1} = (1-\theta)\boldsymbol{U}_i^n + \theta\frac{\boldsymbol{U}_{i-1}^n + \boldsymbol{U}_{i+1}^n}{2} + \frac{\Delta t}{2\Delta x}A\left(\boldsymbol{U}_{i+1}^n - \boldsymbol{U}_{i-1}^n\right)$$

where

$$U_i^n \approx \begin{pmatrix} E(t_n, i\Delta x) \\ H(t_n, i\Delta x) \end{pmatrix} \text{ and } A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Its Fourier symbol matrix is

$$Q_\theta = \Big(1 - \theta + \theta\cos(\tilde{k}h)\Big)\, I + i\lambda\sin(\tilde{k}h)A$$

which satisfies all the conditions in Theorem-1 and 3 in Chapter 2, and a BFECC scheme based on the θ-scheme is second order accurate. To compute the CFL number for the corresponding BFECC scheme, we calculate

$$|\rho(Q_\theta)|^2 = \Big[(1-\theta) + \theta\cos(\tilde{k}h)\Big]^2 + \lambda^2\sin^2(\tilde{k}h)$$

Examine the first term. Since $f(x) = x^2$ is convex, we have

$$\Big[(1-\theta) + \theta\cos(\tilde{k}h)\Big]^2 \le (1-\theta) + \theta\cos^2(\tilde{k}h)$$

Therefore

$$|\rho(Q_\theta)|^2 \le (1-\theta) + \theta\cos^2(\tilde{k}h) + \lambda^2\sin^2(\tilde{k}h) = (1-\theta)|\rho(Q_{\mathcal{L}})|^2 + \theta|\rho(Q_{\mathcal{M}})|^2$$

where $Q_{\mathcal{L}}$ and $Q_{\mathcal{M}}$ are the Fourier symbol matrices for the central difference and Lax-Friedrichs schemes, respectively. Therefore, we have the CFL number of the θ-scheme is

between $\sqrt{3}$ and 2.

Similarly, for two dimensional Maxwell's equations, the θ-scheme is

$$\begin{aligned}\boldsymbol{U}_{i,j}^{n+1} =&(1-\theta)\boldsymbol{U}_{i,j}^{n} + \theta\frac{\boldsymbol{U}_{i-1,j}^{n} + \boldsymbol{U}_{i+1,j}^{n} + \boldsymbol{U}_{i,j-1}^{n} + \boldsymbol{U}_{i,j+1}^{n}}{4} \\ &+ \frac{\Delta t}{2\Delta x}A_1\left(\boldsymbol{U}_{i+1,j}^{n} - \boldsymbol{U}_{i-1,j}^{n}\right) + \frac{\Delta t}{2\Delta y}A_2\left(\boldsymbol{U}_{i,j+1}^{n} - \boldsymbol{U}_{i,j-1}^{n}\right)\end{aligned}$$

where

$$\boldsymbol{U}_{i,j}^{n} \approx \begin{pmatrix} H_x(t_n, i\Delta x, j\Delta y) \\ H_y(t_n, i\Delta x, j\Delta y) \\ E_z(t_n, i\Delta x, j\Delta y) \end{pmatrix},\ A_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},\ A_2 = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}$$

And Fourier symbol matrix

$$Q_\theta = q_\theta I + i\lambda_x \sin(\tilde{k}_x h_x)A_1 + i\lambda_y \sin(\tilde{k}_y h_y)A_2$$

where $q_\theta = \left[1 - \theta + \theta\frac{\cos(\tilde{k}_x h_x)+\cos(\tilde{k}_y h_y)}{2}\right]$, and its spectral radius $\rho(Q_\theta)$ satisfies

$$|\rho(Q_\theta)|^2 = q_\theta^2 + \lambda_x^2 \sin^2(\tilde{k}_x h_x) + \lambda_y^2 \sin^2(\tilde{k}_y h_y) \le (1-\theta)|\rho(Q_{\mathcal{L}})|^2 + \theta|\rho(Q_{\mathcal{M}})|^2$$

In the equality, we again used the convexity of $f(x) = x^2$ and the special form of q_θ. Therefore, the constant in the CFL condition would be between $\sqrt{3}$ and 2. The analysis for three dimensional Maxwell's equations is similar, and the result is summarized as

Theorem 6. *Let $\theta \in [0,1]$, define the θ-scheme $\mathcal{L}_\theta = (1-\theta)\mathcal{L} + \theta\mathcal{M}$, where $\mathcal{L}$ is the central difference scheme and $\mathcal{M}$ is the Lax-Friedrichs scheme for Maxwell's equations. Then the BFECC scheme based on $\mathcal{L}_\theta$ is second order accurate. It is stable if*

1. *in one dimensional case, $\Delta t \le c_\theta \Delta x$,*
2. *in two dimensional case, $\Delta t \le \frac{c_\theta}{\sqrt{(1/\Delta x)^2+(1/\Delta y)^2}}$ and $\Delta t \le \sqrt{\frac{7}{2}}\min(\Delta x, \Delta y)$, and*

3. in three dimensional case,

$$\Delta t \leq \frac{c_\theta}{\sqrt{(1/\Delta x)^2 + (1/\Delta y)^2 + (1/\Delta z)^2}} \text{ and } \Delta t \leq \sqrt{3}\min(\Delta x, \Delta y, \Delta z),$$

where $c_\theta \in [\sqrt{3}, 2]$ is a constant that depends only on θ.

3.6 Least square gradient approximation for non-rectangular grid

In the above discussion of BFECC schemes for the Maxwell's equations, we see that the underlying schemes (central difference or Lax-Friedrichs) are only required to be first order. The BFECC method improves the underlying scheme's accuracy and stability. In order to adapt the scheme to non-orthogonal grid, we consider a simple first order scheme based on least square gradient approximation as the underlying scheme for non-orthogonal or unstructured grid.

To design an explicit scheme for the Maxwell's equations, we need an approximation for the spatial derivatives such as $\frac{\partial E_z}{\partial x}$ and $\frac{\partial H_x}{\partial y}$ at time t_n to update field variables $\boldsymbol{E}$ and $\boldsymbol{H}$. A natural idea is to locally fit a linear function for each component of the field variable using the function values at the a grid point and its neighbors, and then use the fitted coefficients as approximation for the spatial derivatives at this grid point.

We illustrate this idea with the approximation of H_x and its derivatives at a grid point (x_i, y_j). Denote this point (x^0, y^0). Suppose its neighboring grid points are (x^1, y^1), (x^2, y^2), ..., (x^K, y^K), where $K \geq 2$, denote $(H_x)^i = H_x(x^i, y^i)$ for $i = 0, 1, ..., K$. Collect the points and the values $\{(x^i, y^i, (H_x)^i) : i = 0, 1, 2, ..., K\}$, we look a linear function $\hat{H}_x(x, y) = \hat{a} + \hat{b}x + \hat{c}y$ to fit H_x in this neighborhood, i.e. solving the following

(overdetermined) system of linear equations:

$$\begin{pmatrix} 1 & x^0 & y^0 \\ 1 & x^1 & y^1 \\ \cdots & \cdots & \cdots \\ 1 & x^K & y^K \end{pmatrix} \begin{pmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{pmatrix} = \begin{pmatrix} (H_x)^0 \\ (H_x)^1 \\ \cdots \\ (H_x)^K \end{pmatrix} \tag{3.4}$$

We can then use $\hat{b}$ as an approximation for $\frac{\partial H_x}{\partial x}$ and $\hat{c}$ as an approximation for $\frac{\partial H_x}{\partial y}$ at the grid point (x^0, y^0). Such a system is usually over-determined, and we can instead look for the least square solution (the one that minimizes the 2-norm of residual). Note this fitted function $\hat{H}_x$ is only used in a neighborhood of point (x^0, y^0). If we change point, the fitted function will also be changed. We denote the approximated spatial derivatives at (x^0, y^0) by $\frac{\partial \hat{H}_x}{\partial x}$ and $\frac{\partial \hat{H}_x}{\partial y}$, and the approximated function value at (x^0, y^0) by $\hat{H}_x(x^0, y^0)$.

Similarly let $\frac{\partial \hat{E}_z}{\partial x}$, $\frac{\partial \hat{E}_z}{\partial y}$, $\frac{\partial \hat{H}_y}{\partial x}$, $\frac{\partial \hat{H}_y}{\partial y}$ be the least square approximation of E_z and H_y's partial derivatives. An explicit scheme similar to the central difference scheme is then (for Maxwell's equations in two dimensional free space – $\mathrm{TM_z}$ mode):

$$\begin{aligned} (E_z)_{i,j}^{n+1} &= (E_z)_{i,j}^{n} + \Delta t \left(\left(\frac{\partial \hat{H}_y}{\partial x} \right)_{i,j}^{n} - \left(\frac{\partial \hat{H}_x}{\partial y} \right)_{i,j}^{n} \right) \\ (H_x)_{i,j}^{n+1} &= (H_x)_{i,j}^{n} - \Delta t \left(\frac{\partial \hat{E}_z}{\partial y} \right)_{i,j}^{n} \\ (H_y)_{i,j}^{n+1} &= (H_y)_{i,j}^{n} + \Delta t \left(\frac{\partial \hat{E}_z}{\partial x} \right)_{i,j}^{n} \end{aligned} \tag{3.5}$$

It is easy to check that when the grid is a uniform rectangular grid, then the above least square approximation for spatial derivatives is just then central difference approximation, and (3.5) is just the central difference scheme on a uniform rectangular grid. We refer to (3.5) as the least square central difference scheme.

Another explicit scheme based on least square gradient approximation uses the least

square approximated field values as well as the least square approximated derivatives, i.e.

$$
\begin{aligned}
(E_z)_{i,j}^{n+1} &= \left(\hat{E}_z(x^0, y^0)\right)_{i,j}^{n} + \Delta t \left(\left(\frac{\partial \hat{H}_y}{\partial x}\right)_{i,j}^{n} - \left(\frac{\partial \hat{H}_x}{\partial y}\right)_{i,j}^{n} \right) \\
(H_x)_{i,j}^{n+1} &= \left(\hat{H}_x(x^0, y^0)\right)_{i,j}^{n} - \Delta t \left(\frac{\partial \hat{E}_z}{\partial y}\right)_{i,j}^{n} \\
(H_y)_{i,j}^{n+1} &= \left(\hat{H}_y(x^0, y^0)\right)_{i,j}^{n} + \Delta t \left(\frac{\partial \hat{E}_z}{\partial x}\right)_{i,j}^{n}
\end{aligned}
\tag{3.6}
$$

where $\left(\hat{E}_z(x^0, y^0)\right)_{i,j}^{n}$, $\left(\hat{H}_x(x^0, y^0)\right)_{i,j}^{n}$ and $\left(\hat{H}_y(x^0, y^0)\right)_{i,j}^{n}$ are least square approximations to the field values at the grid point with index (i, j). Here the subscript (i, j) and superscript n refer to the fact that this approximation is done in a neighborhood of grid point (x_i, y_j) using field values at time level t_n. Note $\left(\hat{E}_z(x^0, y^0)\right)_{i,j}^{n}$, $\left(\hat{H}_x(x^0, y^0)\right)_{i,j}^{n}$ and $\left(\hat{H}_y(x^0, y^0)\right)_{i,j}^{n}$ are weighted averages of field values at (i, j) and its neighbors, therefore this scheme is similar to the θ-scheme in uniform rectangular grid (note the weights are now not pre-specified but determined by the least square fitting procedure). When the grid is a uniform rectangular grid (possibly with $\Delta x \neq \Delta y$), this reduces to the θ-scheme on uniform rectangular grid with $\theta = 0.8$. We refer to this scheme the least square θ-scheme.

Next, we show that both schemes are first order accurate. To show this, we just need to show the least square gradient approximation are first order accurate, and the least square field value approximation is second order accurate. Without loss of generality, we can assume $(x^0, y^0) = (0, 0)$. In a neighborhood of $(0, 0)$ whose radius is $\Theta(h)$, rewrite function $u(x, y)$ as

$$u(x, y) = a + bx + cy + f(x, y) = l(x, y) + f(x, y)$$

where $f(x, y) = O(x^2 + y^2)$. Suppose the least square fitted function is

$$\hat{u}(x, y) = \hat{a} + \hat{b}x + \hat{c}y$$

To show the least square gradient approximation is first order accurate, we would like to show $||(\hat{a},\hat{b},\hat{c})-(a,b,c)|| \leq O(h)$. Denote $\theta=(a,b,c)^T$ and $\hat{\theta}=(\hat{a},\hat{b},\hat{c})^T$. Suppose (x^0,y^0)'s neighboring grid points are $(x^1,y^1),(x^2,y^2),...,(x^K,y^K)$, where $\sqrt{(x^j)^2+(y^j)^2}=\Theta(h)$, for $j=1,2,...,K$. The coordinates of these points are collected in matrix A as following

$$A=\begin{pmatrix} 1 & x^0 & y^0 \\ 1 & x^1 & y^1 \\ ... & ... & ... \\ 1 & x^K & y^K \end{pmatrix} \tag{3.7}$$

and function values at these grid points are collected in vector U as following

$$U=\begin{pmatrix} l(x^0,y^0)+f(x^0,y^0) \\ l(x^1,y^1)+f(x^1,y^1) \\ ... \\ l(x^K,y^K)+f(x^K,y^K) \end{pmatrix}=L+F$$

where $L=\big(l(x^0,y^0),...,l(x^K,y^K)\big)^T$ and $F=\big(f(x^0,y^0),...,f(x^K,y^K)\big)^T$. Then we have

$$\hat{\theta}=(A^TA)^{-1}A^TU$$
$$\theta=(A^TA)^{-1}A^TL$$

Therefore

$$A(\hat{\theta}-\theta)=A(A^TA)^{-1}A^T(U-L)=A(A^TA)^{-1}A^TF$$
$$\Rightarrow ||A(\hat{\theta}-\theta)||=||A(A^TA)^{-1}A^TF|| \leq ||F||$$

In the above, we use the fact that $A(A^TA)^{-1}A^T$ is an orthogonal projection.

Note A is a $(K+1)\times 3$ matrix of full rank, so its smallest singular value $\sigma_3(A) > 0$. Suppose $\sigma_3(A) \geq Dh$ for some constant $D > 0$, then we have

$$\begin{aligned} &Dh||(\hat{\theta}-\theta)|| \leq \sigma_3||(\hat{\theta}-\theta)|| \leq ||A(\hat{\theta}-\theta)|| \leq ||F|| \leq C(K+1)h^2 \\ \Rightarrow &||(\hat{\theta}-\theta)|| \leq \frac{C(K+1)}{D}h \end{aligned}$$

So the problem reduces to a geometric condition $\sigma_3(A) \geq Dh$ for some $D > 0$ on the grid points. It can be easily verified that the rectangular mesh and the hexagonal mesh both satisfy this condition. For example, an rectangular grid of side length h has $\sigma_3(A) = 2h$, and a hexagonal grid of side length h has $\sigma_3(A) = \sqrt{3}h$.

Next, to show the least square field value approximation is second order accurate, we note the first component of $A(\hat{\theta}-\theta)$ is

$$\hat{a} + \hat{b}x^0 + \hat{c}y^0 - (a + bx^0 + cy^0) = \hat{u}^0 - l(x^0, y^0)$$

where $\hat{u}^0$ is the least square field value approximation, and $l(x^0, y^0) = u(x^0, y^0) - f(x^0, y^0) = u(x^0, y^0)$ since f is the sum of the second and high order terms that all vanish at $(x^0, y^0) = (0, 0)$. Therefore we get:

$$|\hat{u}^0 - u(x^0, y^0)| \leq ||A(\hat{\theta}-\theta)|| \leq C(K+1)h^2$$

So the least square field value approximation is second order accurate.

From the above discussion, we see the least square central difference scheme 3.5 and the least square θ-scheme 3.6 have second order local error, and therefore first order global error.

The order of accuracy result is summarized in the following theorem.

Theorem 7. *Suppose the grid points coordinate matrix A defined in 3.7 satisfies $\sigma_3(A) \geq Dh$ for some positive constant D, then the least square center difference scheme and the*

least square θ-scheme are both first order accurate.

Similar to the central difference scheme, the least square central difference scheme is numerically unstable. We can apply the BFECC method to improve the stability and accuracy. The least square θ-scheme is conditionally stable, and applying BFECC also improves its stability and accuracy. On a uniform rectangular grid, Theorem-1 and 3 in Chapter 2 can be applied to the least square central difference and least square θ-scheme, implying they are second order accurate and stable with CFL number $\sqrt{3}$ and a CFL number between $\sqrt{3}$ and 2, respectively. On non-uniform or non-orthogonal grids, our current analysis is not sufficient to prove the stability and order of accuracy. Numerical examples in Section-3.10 show that BFECC + least square θ-scheme is conditionally stable and second order accurate. We omit the examples for BFECC + least square central difference scheme, it is also second order in our experiments (not reported here) but has larger numerical errors and is less stable.

Remark As will be discussed in section 3.8, on a uniform rectangular grid, central difference scheme and the BFECC scheme based on it preserves the divergence free property of the magnetic field.On a non-rectangular grid, the least square gradient approximation scheme and the corresponding BFECC schemes don't have this property. The flexibility of least square gradient approximation allows an improvement to reduce this error. In addition to the equations 3.4, we can add a penalty term $\lambda \left(\left(\frac{\partial \hat{H}_x}{\partial x} \right) + \left(\frac{\partial \hat{H}_y}{\partial y} \right) \right)^2$ to least square target function, where $\lambda > 0$ is a parameter specifying the weight of the divergence term. This helps reduce the error in the divergence of magnetic field. The Gauss's law for the electric field can similarly be incorporated.

3.7 Point shift algorithm for grid generation

It is often necessary to model curved material interfaces in computational eletromagnetics. The simplest treatment with a staircased approximation for the curved boundary can lead to large errors [27, 6]. Local subcell methods [6] model curved interfaces/boundaries by

modifying the update rule near the interface/boundary. In these cells, the integral form of the Maxwell's equations are usually used to update the field, for example, the contour path method [28].

Using BFECC + least square central difference scheme (3.5) or BFECC + least square θ-scheme (3.6), we can directly deform the grid near a curved interface to conform with the interface, and avoid switching to integral form of the Maxwell's equations in these deformed cells. In this section, we describe a simple algorithm that modifies a uniform rectangular grid locally to conform curved material interfaces. It is used for numerical examples of scattering in Section-3.10.

Given a uniform rectangular grid 2D, denote the grid points $G_{\text{rec}} = \{(x_i, y_j) : x_i = i\Delta x, y_j = j\Delta y, i = 0, 1, ..., N_x, j = 0, 1, ..., N_y\}$, and the grid lines

$$L_{\text{rec}} = \left(\bigcup_{i=0}^{N_x} \{(x, y) : x = x_i, y_0 \leq y \leq y_{N_y}\} \right) \bigcup \left(\bigcup_{j=0}^{N_y} \{(x, y) : y = y_j, x_0 \leq x \leq x_{N_x}\} \right).$$

Let C be a closed curve, for example, the boundary of a scattering object. The point shift algorithm deforms the uniform rectangular grid G_{rec} and generates a new grid $G_C = \{(\tilde{x}_i, \tilde{y}_j) : j = 0, 1, ..., N_y\}$ that conforms with the curves C. It does so by finding the intersection of the grid lines L_{rec} with C, and shifts the nearest grid points to the intersection points. Similar algorithm has been used in interface treatment in two phase flows [29].

Remark A optional smoothing step can be added after the point shift to make the grid deformation more smooth. Denote the uniform rectangular grid points $\boldsymbol{x}_{i,j}$ and the point shifted grid point $\tilde{\boldsymbol{x}}_{i,j}$, where $i = 0, 1, ..., N_x$ and $j = 0, 1, ..., N_y$. First compute the point shift deformation $\boldsymbol{d}_{i,j} = \tilde{\boldsymbol{x}}_{i,j} - \boldsymbol{x}_{i,j}$. Second, copy $\boldsymbol{d}_{i,j}$ to $\tilde{\boldsymbol{d}}_{i,j}$, and for every (i, j) such that $\boldsymbol{d}_{i,j} = \boldsymbol{0}$ (i.e. unshifted points), set

$$\tilde{\boldsymbol{d}}_{i,j} = \frac{\boldsymbol{d}_{i-1,j} + \boldsymbol{d}_{i+1,j} + \boldsymbol{d}_{i,j-1} + \boldsymbol{d}_{i,j+1}}{4}.$$

This has the effect of smoothing out the point shift deformation. Third, assign new locations

Algorithm 1: Point shift algorithm

function PointShift $(G_{\text{rec}}, L_{\text{rec}}, C)$;

Input : Rectangular grid
$G_{\text{rec}} = \{(x_i, y_j) : x_i = i\Delta x, y_j = j\Delta y, i = 0, 1, ..., N_x, j = 0, 1, ..., N_y\}$,
the grid lines L_{rec}, curve C.

Output: Deformed grid $G_C = \{(\tilde{x}_i, \tilde{y}_j) : j = 0, 1, ..., N_y\}$

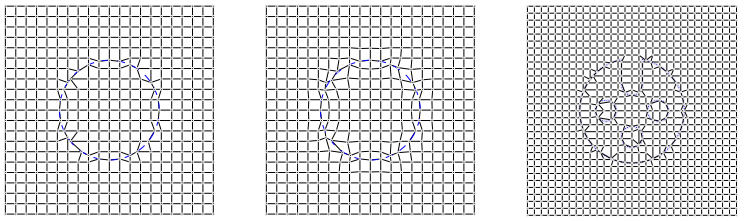
1. Copy G_{rec} to G_C: set $\tilde{x}_i = x_i$, $\tilde{y}_j = x_j$ for $i = 0, 1, ..., N_x$, $j = 0, 1, ..., N_y$;
2. Find intersections points $\{(\hat{x}^k, \hat{y}^k) : k = 0, 1, ..., K\}$ of L_{rec} and C;
3. **for** *k = 0, 1, ..., K* **do**
   Find the nearest point $(\tilde{x}_{i^*}, \tilde{y}_{j^*})$ in G_C to $(\hat{x}^k, \hat{y}^k)$, when there is a tie, break the tie arbitrarily. Set $(\tilde{x}_{i^*}, \tilde{y}_{j^*}) = (\hat{x}^k, \hat{y}^k)$.
   end
4. Return G_C.

to the shifted grid points

$$\tilde{\boldsymbol{x}}_{i,j} = \boldsymbol{x}_{i,j} + \tilde{\boldsymbol{d}}_{i,j}.$$

for $i = 0, 1, ..., N_x$ and $j = 0, 1, ..., N_y$. Note the shifted grid points that lie on the curve C are unaffected by this smoothing step, only their neighbors get shifted in the smoothing step. This step can be repeated multiple times to smooth out the deformation to points that are further away from the curve C. Smoothing helps reduce grid deformation near the interface, and can be helpful when complicated interfaces are involved.

Figure-3.1 shows examples of non-rectangular grids obtained from point shift. The subfigure (a) is a uniform rectangular grid shifted to conform a circle without smoothing, the subfigure (b) is the same grid shifted to conform a circle, with a smoothing step, and the subfigure (c) is a uniform rectangular grid shifted to conform a more complicated curve, without smoothing. Grid (a) and (c) are use in the scattering numerical examples in Section-3.10. We didn't use the smoothing step since the material interfaces in our numerical examples are simple and solutions on grids without smoothing already has expected order of accuracy.



(a) Point shifted to conform a circular interface.

(b) Point shifted to conform a circular interface, followed by a smoothing step.

(c) Point shifted to conform a complicated interface.

Figure 3.1: Point shifted grids.

3.8 Divergence of magnetic field

The magnetic field satisfies the divergence free condition in Maxwells' equation. We show that the central difference scheme preserves the numerical divergence of the magnetic field when the grid is a uniform rectangular grid. Therefore, the BFECC + central difference also preserves the numerical divergence of the magnetic field.

The numerical divergence of the magnetic field at time step n is:

$$(\nabla \cdot \vec{H})^n_{i,j} = \frac{(H_x)^n_{i+1,j} - (H_x)^n_{i-1,j}}{2\Delta x} + \frac{(H_y)^n_{i,j+1} - (H_y)^n_{i,j-1}}{2\Delta y}$$

Using the central difference scheme to update H_x and H_y, we get:

$$\begin{aligned}
\frac{(H_x)^{n+1}_{i+1,j} - (H_x)^{n+1}_{i-1,j}}{2\Delta x} =& \frac{(H_x)^n_{i+1,j} - (H_x)^n_{i-1,j}}{2\Delta x} - \\
& \frac{(E_z)^n_{i+1,j+1} - (E_z)^n_{i+1,j-1} - (E_z)^n_{i-1,j+1} + (E_z)^n_{i-1,j-1}}{4\Delta x \Delta y}\Delta t \\
\frac{(H_y)^{n+1}_{i,j+1} - (H_y)^{n+1}_{i,j-1}}{2\Delta x} =& \frac{(H_y)^n_{i,j+1} - (H_y)^n_{i,j-1}}{2\Delta y} + \\
& \frac{(E_z)^n_{i+1,j+1} - (E_z)^n_{i+1,j-1} - (E_z)^n_{i-1,j+1} + (E_z)^n_{i-1,j-1}}{4\Delta x \Delta y}\Delta t
\end{aligned}$$

Therefore

$$(\nabla \cdot \boldsymbol{H})_{i,j}^{n+1} = (\nabla \cdot \boldsymbol{H})_{i,j}^{n}$$

Similar argument shows that the Lax-Friedrichs scheme and the θ-scheme preserve $\nabla \cdot \boldsymbol{H}$.

For irregular grid, the divergence free property is no longer guaranteed. But we can add divergence penalty terms during the least square gradient approximation to reduce this error, as discussed in section 3.6.

3.9 Perfectly Matched Layer

In applications of computational electromagnetics, it is often necessary to simulate wave propagation in unbounded domains. Computationally, this usually translates to a boundary condition that allows waves to propagate out of the computation domain freely and prohibits waves to propagate into the computation domain. These boundary conditions are referred to as the absorbing boundary conditions (ABCs).

The first class of commonly used absorbing boundary conditions are obtained by approximating the exact differential-integral boundary condition with differential boundary conditions [30, 6]. The second class of absorbing boundary conditions are actually absorbing boundary layers that absorbs the out-going waves, and they are known as the perfectly matched layers (PMLs) [19].

For completeness, we first review the unsplit convolutional perfectly matched layers [31] (unsplit CPML) in this section, following the treatment in [32] for PMLs for wave equations. Then we propose an implementation with the BFECC method and present numerical examples to verify the effectiveness of this implementation. Our contribution is the new BFECC implementation for the unsplit CPML.

3.9.1 One dimensional case

Consider solving Maxwell's equations in one dimension,

$$\frac{\partial H_y}{\partial t} = \frac{\partial E_z}{\partial x}$$
$$\frac{\partial E_z}{\partial t} = \frac{\partial H_y}{\partial x}$$

with some initial conditions. To simplify notation, denote $E = E_z, H = H_y$ and we have:

$$\frac{\partial H}{\partial t} = \frac{\partial E}{\partial x}$$
$$\frac{\partial E}{\partial t} = \frac{\partial H}{\partial x}$$

Suppose our computation domain $\{x : x \leq 0\}$, we need to impose some boundary condition at $x = 0$ or design some absorbing layers in $x \geq 0$ so that electromagnetic waves can freely propagate from the computation domain to $\{x : x \geq 0\}$.

Remark Note here we could allow the electric current term J_z to be nonzero in the computational domain.

Consider the equations in $\{x : x \geq 0\}$:

$$\frac{\partial H}{\partial t} = \frac{\partial E}{\partial x}$$
$$\frac{\partial E}{\partial t} = \frac{\partial H}{\partial x}, \; x > 0$$

with $E(0, x) = 0, H(0, x) = 0$ and $E(0, t) = E(t), H(0, t) = H(t)$, where $E(t)$ and $H(t)$ are the solutions from the whole space Maxwell's equations.

First we extend $E(t, x)$ and $H(t, x)$ to the region $x < 0$. The extended $\tilde{E}(t, x)$ and $\tilde{H}(t, x)$ still satisfies the same set of equations, with $\tilde{E}(0, x) = 0$ and $\tilde{H}(0, x) = 0$. With some abuse of notation, we still denote $\tilde{E}$ by E and $\tilde{H}$ by H.

Apply Fourier transform (in t) to the equations, we get:

$$\frac{\partial \hat{E}}{\partial x} = i\omega \hat{H}$$
$$\frac{\partial \hat{H}}{\partial x} = i\omega \hat{E}$$

Solve this set of ODEs, we get

$$\hat{E}(x) = a_+(\omega)e^{i\omega x} + a_-(\omega)e^{-i\omega x}$$
$$\hat{H}(x) = b_+(\omega)e^{i\omega x} + b_-(\omega)e^{-i\omega x}$$

Apply inverse Fourier transform, we get

$$E(t,x) = \int_{\mathbb{R}} a_+(\omega)e^{i\omega x}e^{i\omega t}d\omega + \int_{\mathbb{R}} a_-(\omega)e^{-i\omega x}e^{i\omega t}d\omega$$
$$H(t,x) = \int_{\mathbb{R}} b_+(\omega)e^{i\omega x}e^{i\omega t}d\omega + \int_{\mathbb{R}} b_-(\omega)e^{-i\omega x}e^{i\omega t}d\omega$$

Consider the $E(t,x)$ and $H(t,x)$ in $x \geq 0$, in the above expressions, the terms with $a_+(\omega)$ and $b_+(\omega)$ are propagating waves from infinity to the origin. Since the only source is located at $x = 0$, these waves should be disregarded. The terms with $a_-(\omega)$ and $b_-(\omega)$ are propagating waves from the origin to infinity. In order to have an absorbing layers in $x > 0$, we need to modify these terms so that they became evanescent waves (i.e. the altitude of the wave decades as it ravels to the infinity). To achieve this goal, first note E and H are analytical functions in x from the above expressions. Therefore we can analytically

continue the E and H to the whole complex plane

$$E(t,z) = \int_{\mathbb{R}} a_-(\omega) e^{-i\omega z} e^{i\omega t} d\omega$$
$$H(t,z) = \int_{\mathbb{R}} b_-(\omega) e^{-i\omega z} e^{i\omega t} d\omega$$

After this analytical continuation, E and H still satisfy the Maxwell's equations, and now we have:

$$\frac{\partial \hat{E}}{\partial z} = i\omega \hat{H}$$
$$\frac{\partial \hat{H}}{\partial z} = i\omega \hat{E}$$

Second, we consider a complex coordinate stretch $S : \mathbb{R} \to \mathbb{C}$ defined as

$$S(x) = x + \frac{1}{i\omega} \int_0^x \sigma(\tau) d\tau \tag{3.8}$$

where $\sigma(\tau) > 0$ is a real function. And define

$$\hat{U}(x) = \hat{E}(S(x))$$
$$\hat{V}(x) = \hat{H}(S(x))$$

for $x > 0$ and $\hat{U}(x) = \hat{E}(x), \hat{V}(x) = \hat{H}(x)$ for $x < 0$.

Let $U = \mathcal{F}^{-1}(\hat{U})$ and $V = \mathcal{F}^{-1}(\hat{V})$, we get: for $x > 0$,

$$U(t,x) = \int_{\mathbb{R}} a_-(\omega) e^{-\int_0^x \sigma(\tau) d\tau} e^{-i\omega x} e^{i\omega t} d\omega$$
$$V(t,x) = \int_{\mathbb{R}} b_-(\omega) e^{-\int_0^x \sigma(\tau) d\tau} e^{-i\omega x} e^{i\omega t} d\omega$$

with $U(t,0+) = E(t,0)$ and $V(t,0+) = H(t,0)$. Note U and V are now evanescent waves.

To derive the differential equations satisfied by U and V, note from 3.8, we have

$$S'(x) = 1 + \frac{\sigma}{i\omega}$$

Therefore

$$\frac{\partial \hat{U}}{\partial x} = \frac{\partial \hat{E}}{\partial z} S'(x) = \left(1 + \frac{\sigma}{i\omega}\right) \frac{\partial \hat{E}}{\partial z}(S(x)) = (\sigma + i\omega)\hat{H}(S(x)) = (\sigma + i\omega)\hat{V}$$
$$\frac{\partial \hat{V}}{\partial x} = \frac{\partial \hat{H}}{\partial z} S'(x) = \left(1 + \frac{\sigma}{i\omega}\right) \frac{\partial \hat{H}}{\partial z}(S(x)) = (\sigma + i\omega)\hat{E}(S(x)) = (\sigma + i\omega)\hat{U}$$

Apply inverse Fourier transform and we get

$$\frac{\partial U}{\partial t} + \sigma U = \frac{\partial V}{\partial x}$$
$$\frac{\partial V}{\partial t} + \sigma V = \frac{\partial U}{\partial x},\ x > 0$$

These are the PML equations for the one dimensional case

$$\frac{\partial E}{\partial t} + \sigma E = \frac{\partial H}{\partial x}$$
$$\frac{\partial H}{\partial t} + \sigma H = \frac{\partial E}{\partial x},\ x > 0$$

3.9.2 Two dimensional case

Consider the two dimensional TM_z mode. The computational domain is $\{x : x \leq 0\}$, and we design proper conditions for absorbing layers $\{x : x \geq 0\}$ so that electromagnatic waves can freely propagate from the computation domain to the absorbing layers.

Starting from Maxwell's equations in $\{x : x \geq 0\}$

$$\begin{aligned}
\frac{\partial H_x}{\partial t} &= -\frac{\partial E_z}{\partial y} \\
\frac{\partial H_y}{\partial t} &= \frac{\partial E_z}{\partial x} \\
\frac{\partial E_z}{\partial t} &= \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}
\end{aligned}$$

with $H_x(0, x, y) = H_y(0, x, y) = 0, E_z(0, x, y) = 0$ and $H_x(t, 0, y) = H_x(t, y), H_y(t, 0, y) = H_y(t, y), E_z(t, 0, y) = E_z(t, y)$, where $H_x(t, y), H_y(t, y)$ and $E_z(t, y)$ are solutions from the whole space Maxwell's equation at $x = 0$.

We can extend H_x, H_y and E_z to the region $x < 0$ so that the extended functions still satisfies Maxwell's equations and initial conditions $H_x(0, x, y) = H_y(0, x, y) = 0$ and $E_z(0, x, y) = 0$. With some abuse of notation, we denote the extended functions by H_x, H_y and E_z.

Apply Fourier transform in t and y to the Maxwell's equations, we get

$$\begin{aligned}
i\omega \tilde{H}_x &= -ik\tilde{E}_z \\
i\omega \tilde{H}_y &= \frac{\partial \tilde{E}_z}{\partial x} \\
i\omega \tilde{E}_z &= \frac{\partial \tilde{H}_y}{\partial x} - ik\tilde{H}_x
\end{aligned}$$

where $\tilde{H}_x$, $\tilde{H}_y$ and $\tilde{E}_z$ are Fourier transform of H_x, H_y and E_z, ω is the frequency variable for t and k is the frequency variable for y. Note this gives a ODE system for $\tilde{E}_z$ and $\tilde{H}_y$ as follows

$$\frac{\partial}{\partial x}\begin{pmatrix} \tilde{E}_z \\ \tilde{H}_y \end{pmatrix} = \begin{pmatrix} 0 & i\omega \\ i\omega\left(1 - \frac{k^2}{\omega^2}\right) & 0 \end{pmatrix}\begin{pmatrix} \tilde{E}_z \\ \tilde{H}_y \end{pmatrix}$$

which has eigenvalues

$$\lambda_+ = \sqrt{k^2-\omega^2},\ \lambda_- = -\sqrt{k^2-\omega^2},\ \text{ if } |k| > |\omega|$$
$$\lambda_+ = i\sqrt{\omega^2-k^2},\ \lambda_- = -i\sqrt{\omega^2-k^2},\ \text{ else.}$$

Therefore, the solution for E_z is

$$\begin{aligned}
E_z(t,x,y) = &\iint_{|k|>|\omega|} a_+(\omega,k)e^{\sqrt{k^2-\omega^2}x}e^{i(\omega t+ky)}d\omega dk + \\
&\iint_{|k|>|\omega|} a_-(\omega,k)e^{-\sqrt{k^2-\omega^2}x}e^{i(\omega t+ky)}d\omega dk + \\
&\iint_{|k|<\omega} a_+(\omega,k)e^{i\sqrt{\omega^2-k^2}x}e^{i(\omega t+ky)}d\omega dk + \\
&\iint_{\omega<-|k|} a_+(\omega,k)e^{i\sqrt{\omega^2-k^2}x}e^{i(\omega t+ky)}d\omega dk + \\
&\iint_{|k|<\omega} a_-(\omega,k)e^{-i\sqrt{\omega^2-k^2}x}e^{i(\omega t+ky)}d\omega dk + \\
&\iint_{\omega<-|k|} a_-(\omega,k)e^{-i\sqrt{\omega^2-k^2}x}e^{i(\omega t+ky)}d\omega dk
\end{aligned}$$

The first term on the right is ruled out by the finiteness so E_z at infinity, and the third and sixth terms are disregarded since they represent waves propagating from positive infinity to the left and the only source term is at $x = 0$. The second term is an evanescent wave. In order to have absorbing layers in $x > 0$, we need to modify the solution so that the fourth and fifth terms become evanescent waves. Similar to the one dimensional case, we first analytically continue E_z (and H_x, H_y) to the whole complex plane

$$\begin{aligned}
E_z(t,z,y) = &\iint_{|k|>|\omega|} a_-(\omega,k)e^{-\sqrt{k^2-\omega^2}z}e^{i(\omega t+ky)}d\omega dk + \\
&\iint_{\omega<-|k|} a_+(\omega,k)e^{i\sqrt{\omega^2-k^2}z}e^{i(\omega t+ky)}d\omega dk + \\
&\iint_{|k|<\omega} a_-(\omega,k)e^{-i\sqrt{\omega^2-k^2}z}e^{i(\omega t+ky)}d\omega dk
\end{aligned}$$

After the continuation, E_z, H_x and H_y satisfies

$$\frac{\partial E_z}{\partial t} = \frac{\partial H_y}{\partial z} - \frac{\partial H_x}{\partial y}$$

Note there z is a complex variable (continued from x), not the third spatial dimension.

Similar to the one dimensional case, we introduce a complex coordinate stretch S : $\mathbb{R} \to \mathbb{C}$:

$$S(x) = x + \frac{1}{i\omega} \int_0^x \sigma(\tau) d\tau \tag{3.9}$$

where $\sigma(\tau) > 0$ is a real function. And define

$$\tilde{U}(\omega, x, k) = \tilde{E}_z(\omega, S(x), k)$$
$$\tilde{V}(\omega, x, k) = \tilde{H}_x(\omega, S(x), k)$$
$$\tilde{W}(\omega, x, k) = \tilde{H}_y(\omega, S(x), k)$$

for $x > 0$ and $\tilde{U}(\omega, x, k) = \tilde{E}_z(\omega, x, k)$, $\tilde{V}(\omega, x, k) = \tilde{H}_x(\omega, x, k)$, $\tilde{W}(\omega, x, k) = \tilde{H}_y(\omega, x, k)$ on $x \leq 0$.

Let $U(t, x, y) = \mathcal{F}_t^{-1}(\mathcal{F}_y^{-1}(\tilde{U}))$, $v(t, x, y) = \mathcal{F}_t^{-1}(\mathcal{F}_y^{-1}(\tilde{V}))$ and $W(t, x, y) = \mathcal{F}_t^{-1}(\mathcal{F}_y^{-1}(\tilde{W}))$, and check the expression for $U(t, x, y)$

$$\begin{aligned} U(t, x, y) = &\iint_{|k|>|\omega|} a_-(\omega, k) e^{-\sqrt{k^2-\omega^2}x} \exp\left(-i\omega \frac{\sqrt{k^2-\omega^2}}{\omega^2} \int_0^x \sigma(\tau) d\tau\right) e^{i(\omega t + ky)} d\omega dk + \\ &\iint_{\omega<-|k|} a_+(\omega, k) e^{i\sqrt{\omega^2-k^2}x} \exp\left(\frac{\sqrt{\omega^2-k^2}}{\omega} \int_0^x \sigma(\tau) d\tau\right) e^{i(\omega t + ky)} d\omega dk + \\ &\iint_{|k|<\omega} a_-(\omega, k) e^{-i\sqrt{\omega^2-k^2}x} \exp\left(-\frac{\sqrt{\omega^2-k^2}}{\omega} \int_0^x \sigma(\tau) d\tau\right) e^{i(\omega t + ky)} d\omega dk \end{aligned}$$

Note the first term is still evanescent (although an oscillatory factor is introduced), and the second and third terms also become evanescent.

Similar to the one dimensional case, we can obtain the equations for $\tilde{U}$, $\tilde{V}$ and $\tilde{W}$

$$
\begin{aligned}
i\omega\tilde{V} &= -ik\tilde{U} \\
i\omega\tilde{W} &= \frac{1}{1+\frac{\sigma}{i\omega}}\frac{\partial\tilde{U}}{\partial x} \\
i\omega\tilde{U} &= \frac{1}{1+\frac{\sigma}{i\omega}}\frac{\partial\tilde{W}}{\partial x} - ik\tilde{V}
\end{aligned}
$$

Suppose σ is a constant, then

$$
\mathcal{F}_t^{-1}\left(\frac{1}{1+\frac{\sigma}{i\omega}}\right) = \delta(t) - \sigma e^{-\sigma t}u(t)
$$

where $\delta(t)$ is the Dirac delta function, and $u(t)$ is the unit step function. And the equations for U, V and W are

$$
\begin{aligned}
\frac{\partial U}{\partial t} &= \frac{\partial W}{\partial x} - \frac{\partial V}{\partial y} - \sigma e^{-\sigma t}u(t) * \frac{\partial W}{\partial x} \\
\frac{\partial V}{\partial t} &= -\frac{\partial U}{\partial y} \\
\frac{\partial W}{\partial t} &= \frac{\partial U}{\partial x} - \sigma e^{-\sigma t}u(t) * \frac{\partial U}{\partial x}
\end{aligned}
$$

where $*$ is the convolution operation.

In order to absorb waves propagating in the y direction, we can apply a similar complex coordinate stretch $S(y) = 1 + \frac{1}{i\omega}\int_0^y \sigma_y(\tau)d\tau$, and transform the equation correspondingly. In summary, with the unsplit convolutional perfectly match layers, the equations in the perfectly matched layers are:

$$
\begin{aligned}
\frac{\partial E_z}{\partial t} &= \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} + \zeta_x(t) * \frac{\partial H_y}{\partial x} - \zeta_y(t) * \frac{\partial H_x}{\partial y} \\
\frac{\partial H_x}{\partial t} &= -\frac{\partial E_z}{\partial y} - \zeta_y(t) * \frac{\partial E_z}{\partial y} \\
\frac{\partial H_y}{\partial t} &= \frac{\partial E_z}{\partial x} + \zeta_x(t) * \frac{\partial E_z}{\partial x}
\end{aligned}
$$

where

$$\zeta_w(t) = -\sigma_w e^{-\sigma_w t} u(t),\ w = x, y,$$

$u(t)$ is the unit step function, and σ_x, σ_y are chosen conductivity parameters in the perfectly matched layers (PMLs). For PMLs adjacent to a boundary perpendicular to the x-axis, we choose $\sigma_x > 0$ and $\sigma_y = 0$, and for PMLs adjacent to a boundary perpendicular to the y-axis, we choose $\sigma_y > 0$ and $\sigma_x = 0$.

3.9.3 Implementation

We next consider the implementation of the un-split convolutional perfectly matched layer [31] with the BFECC method. Here we adapt the implementation in [33] and discuss in the two dimensional case. The three dimensional case is similar.

In the lossless domain, we have:

$$\begin{aligned}
\frac{\partial E_z}{\partial t} &= \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} \\
\frac{\partial H_x}{\partial t} &= -\frac{\partial E_z}{\partial y} \\
\frac{\partial H_y}{\partial t} &= \frac{\partial E_z}{\partial x}
\end{aligned}$$

With the unsplit convolutional perfectly match layers, the equations in the perfectly matched layers are:

$$\begin{aligned}
\frac{\partial E_z}{\partial t} &= \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} + \zeta_x(t) * \frac{\partial H_y}{\partial x} - \zeta_y(t) * \frac{\partial H_x}{\partial y} \\
\frac{\partial H_x}{\partial t} &= -\frac{\partial E_z}{\partial y} - \zeta_y(t) * \frac{\partial E_z}{\partial y} \\
\frac{\partial H_y}{\partial t} &= \frac{\partial E_z}{\partial x} + \zeta_x(t) * \frac{\partial E_z}{\partial x}
\end{aligned}$$

where

$$\zeta_w(t) = -\sigma_w e^{-\sigma_w t} u(t),\ w = x, y,$$

$u(t)$ is the unit step function, and σ_x, σ_y are chosen conductivity parameters in the perfectly matched layers (PMLs). For PMLs adjacent to a boundary perpendicular to the x-axis, we choose $\sigma_x > 0$ and $\sigma_y = 0$, and for PMLs adjacent to a boundary perpendicular to the y-axis, we choose $\sigma_y > 0$ and $\sigma_x = 0$.

To implement BFECC in the perfectly matched layers, we first denote

$$
\begin{aligned}
&b_x = e^{-\sigma_x \Delta t},\ b_y = e^{-\sigma_y \Delta t} \\
&c_x = b_x - 1,\ c_y = b_y - 1 \\
&\left(\Psi_{E_z x}\right)_{i,j}^{n} = \left(\zeta_x(t) * \frac{\partial H_y}{\partial x}\right)_{i,j}^{n} \\
&\left(\Psi_{E_z y}\right)_{i,j}^{n} = \left(\zeta_y(t) * \frac{\partial H_x}{\partial y}\right)_{i,j}^{n} \\
&\left(\Psi_{H_x y}\right)_{i,j}^{n} = \left(\zeta_y(t) * \frac{\partial E_z}{\partial y}\right)_{i,j}^{n} \\
&\left(\Psi_{H_y x}\right)_{i,j}^{n} = \left(\zeta_y(t) * \frac{\partial E_z}{\partial x}\right)_{i,j}^{n}
\end{aligned}
$$

To update the field variables in the PMLs, consider the least square central difference scheme:

$$
\begin{aligned}
\left(E_z\right)_{i,j}^{n+1} = \left(E_z\right)_{i,j}^{n} + \Delta t &\left(\left(\frac{\partial \hat{H}_y}{\partial x}\right)_{i,j}^{n} - \left(\frac{\partial \hat{H}_x}{\partial y}\right)_{i,j}^{n}\right) \\
&+ \left(c_x \left(\frac{\partial \hat{H}_y}{\partial x}\right)_{i,j}^{n} + b_x \left(\Psi_{E_z x}\right)_{i,j}^{n-1}\right) \Delta t \\
&- \left(c_y \left(\frac{\partial \hat{H}_x}{\partial y}\right)_{i,j}^{n} + b_y \left(\Psi_{E_z y}\right)_{i,j}^{n-1}\right) \Delta t
\end{aligned}
$$

$$(H_x)_{i,j}^{n+1} = (H_x)_{i,j}^{n} - \Delta t \left(\frac{\partial \hat{E}_z}{\partial y} \right)_{i,j}^{n} - \left(c_y \left(\frac{\partial \hat{E}_z}{\partial y} \right)_{i,j}^{n} + b_y \left(\Psi_{H_x y} \right)_{i,j}^{n-1} \right) \Delta t$$

$$(H_y)_{i,j}^{n+1} = (H_y)_{i,j}^{n} + \Delta t \left(\frac{\partial \hat{E}_z}{\partial x} \right)_{i,j}^{n} + \left(c_x \left(\frac{\partial \hat{E}_z}{\partial x} \right)_{i,j}^{n} + b_x \left(\Psi_{H_y x} \right)_{i,j}^{n-1} \right) \Delta t$$

Here $\hat{H}_x$, $\hat{H}_y$ and $\hat{E}_z$ are the least square reconstructed functions.

To apply the BFECC method to this scheme, we combine all the terms on the right hand side that involve spatial derivatives, for example, the equation for E_z becomes

$$(E_z)_{i,j}^{n+1} = (E_z)_{i,j}^{n} + \Delta t \left((1 + c_x) \left(\frac{\partial \hat{H}_y}{\partial x} \right)_{i,j}^{n} - (1 + c_y) \left(\frac{\partial \hat{H}_x}{\partial y} \right)_{i,j}^{n} \right) + \left(b_x \left(\Psi_{E_z x} \right)_{i,j}^{n-1} - b_y \left(\Psi_{E_z y} \right)_{i,j}^{n-1} \right) \Delta t$$

The $\left(b_x \left(\Psi_{E_z x} \right)_{i,j}^{n-1} - b_y \left(\Psi_{E_z y} \right)_{i,j}^{n-1} \right)$ term is treated as a source term. In the first two steps of the BFECC method, we ignore this source term. It is only added in the third step of BFECC method. We see this requires very little modification to the update rule in the computation domain.

Similarly, we can also use BFECC + least square θ-scheme in the PMLs.

The above equations updates the field variables, with the updated field variables, we

update the convolutional quantities

$$
\begin{aligned}
\left(\Psi_{E_z x}\right)_{i,j}^{n} &= c_x \left(\frac{\partial \hat{H}_y}{\partial x}\right)_{i,j}^{n} + b_x \left(\Psi_{E_z x}\right)_{i,j}^{n-1} \\
\left(\Psi_{E_z y}\right)_{i,j}^{n} &= c_y \left(\frac{\partial \hat{H}_x}{\partial y}\right)_{i,j}^{n} + b_y \left(\Psi_{E_z y}\right)_{i,j}^{n-1} \\
\left(\Psi_{H_x y}\right)_{i,j}^{n} &= c_y \left(\frac{\partial \hat{E}_z}{\partial y}\right)_{i,j}^{n} + b_y \left(\Psi_{H_x y}\right)_{i,j}^{n-1} \\
\left(\Psi_{H_y x}\right)_{i,j}^{n} &= c_x \left(\frac{\partial \hat{E}_z}{\partial x}\right)_{i,j}^{n} + b_x \left(\Psi_{H_y x}\right)_{i,j}^{n-1}
\end{aligned}
$$

3.10 Numerical examples

3.10.1 1D periodic solution

We consider the following periodic initial condition for the 1D Maxwell's equations

$$E(0, x) = H(0, x) = \sin(2\pi x)$$

The solution that satisfies the given initial condition is

$$E(t, x) = H(t, x) = \sin 2\pi(x + t)$$

We solve the system with FDTD Yee scheme and BFECC scheme based on central difference from $t = 0$ to $t = 0.6$ with $\Delta t/\Delta x = 0.38$, 0.98 and 1.5, and compare the numerical solutions with the exact solution.

The order of accuracy result is summarized in Table-3.1. The results confirms BFECC + central difference scheme is second order accurate. Also note the scheme is stable for $\Delta t = 1.5\Delta x$, , for which the classical Yee scheme becomes unstable.

Table 3.1: Order of accuracy for BFECC + central difference scheme at $T = 0.6$

Grid	$\Delta t/\Delta x = 0.38$		$\Delta t/\Delta x = 0.98$		$\Delta t/\Delta x = 1.5$	
	Error	Order	Error	Order	Error	Order
64	0.0111	–	0.0250	–	0.0456	–
128	0.0028	2.00	0.0064	1.97	0.0115	1.99
256	7.9306×10^{-4}	2.00	0.0016	2.00	0.0029	1.98
512	1.7328×10^{-4}	2.00	4.0003×10^{-4}	2.00	7.3476×10^{-4}	1.99
1024	4.3320×10^{-5}	2.00	1.0023×10^{-4}	2.00	1.8370×10^{-4}	2.00
2048	1.0830×10^{-5}	2.00	2.5085×10^{-5}	2.00	4.5926×10^{-5}	2.00

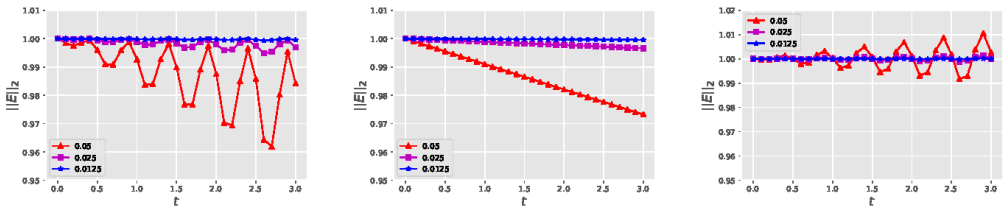
3.10.2 Comparison of BFECC central difference, Lax-Friedrichs, and θ-schemes

We compare the BFECC + central difference, BFECC + Lax-Friedrichs, and BFECC + θ-schemes for their error order, dissipation and dispersion errors in this example. The Maxwell's equation in $\mathrm{TM_z}$ mode is solved with periodic boundary condition and plane wave initial data $E_z(0, x, y) = \sin(2\pi x)$, $H_x(0, x, y) = 0$ and $H_y(0, x, y) = -\sin(2\pi x)$. The ratio $\Delta t/\Delta x$ is fixed to be 0.5. The numerical error and order of accuracy for solutions at $T = 0.25$ is shown in Table-3.2. We see all three schemes are at least second order accurate (the BFECC + Lax-Friedrichs scheme actually has a numerical accuracy of the third order). The error of BFECC + Lax-Friedrichs scheme is also significantly smaller than BFECC + central difference scheme. θ-scheme, as a combination of the central difference (CD) and Lax-Friedrichs schemes (LF), has performance between CD and LF.

Table 3.2: Error and order of accuracy for BFECC + central difference(CD), Lax-Friedrichs(LF), and θ-scheme ($\theta = 0.5$) at $T = 2.5$

Grid	BFECC + CD		BFECC + LF		BFECC + θ-scheme ($\theta = 0.5$)	
	Error	Order	Error	Order	Error	Order
20×20	0.2825	–	1.622×10^{-2}	–	1.452×10^{-1}	–
40×40	0.0705	2.00	2.036×10^{3}	2.99	3.543×10^{-2}	2.04
80×80	0.0174	2.02	2.536×10^{-4}	3.01	8.718×10^{-3}	2.03

To have a better understanding of the errors of these three scheme, we look at the energy dissipation and numerical wave propagation speed. An ideal numerical scheme preserves energy of the field, thus its Fourier symbol matrix is unitary; it also propagates waves of any frequency at the correct speed ($c = 1$ in our normalized equations). Practical schemes violate one or both of these conditions. We first compare the energy dissipation of the three schemes. As shown in Figure-3.2 , BFECC + CD dissipates energy in a oscillatory fashion, and BFECC + LF dissipates energy linearly with time, the rate of which is slightly larger than the average dissipation rate of BFECC + CD. BFECC + θ-scheme seems to be better than both, with a smaller oscillation and on average preserves the electric field energy very well.



(a) BFECC + central difference: $\Delta x = 0.05, 0.025, 0.0125$. (b) BFECC + Lax-Friedrichs: $\Delta x = 0.05, 0.025, 0.0125$. (c) BFECC + θ-scheme: $\Delta x = 0.05, 0.025, 0.0125$.

Figure 3.2: Comparison of numerical dissipation.

Figure-3.3 compares the solution profiles on $y = 0.5$ slice at $t = 2.5$. We see BFECC +

CD solution has a slower than 1 numerical wave speed, BFECC + LF has numerical wave speed very close to 1, and BFECC + θ-scheme again sits in between.

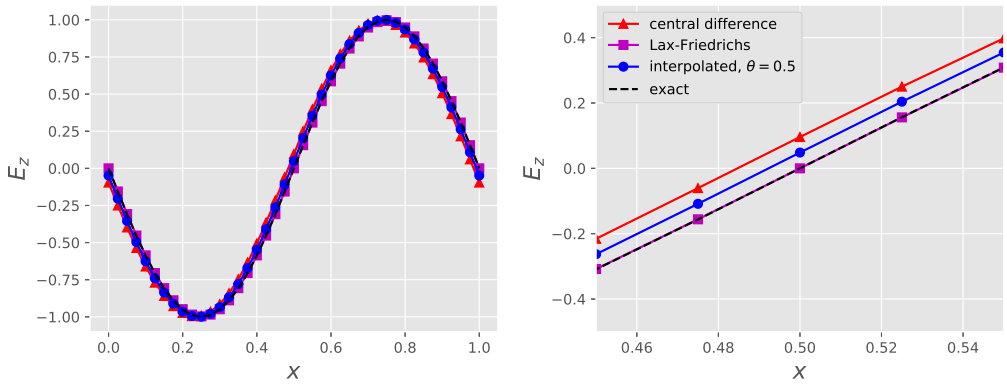


Figure 3.3: Comparison of numerical propagation speed of plane wave.

In summary, BFECC + Lax-Friedrichs scheme has the smallest numerical error but is the most dissipative one. BFECC + central difference scheme has the largest numerical error, but is less dissipative. BFECC + θ-scheme's error is in between but has the best energy conservation performance.

3.10.3 2D periodic solution

We consider the following periodic initial condition for the 2D Maxwell's equations in $\mathrm{TM_z}$ mode.

$$E_z(0, x, y) = \sin(2\pi x)$$

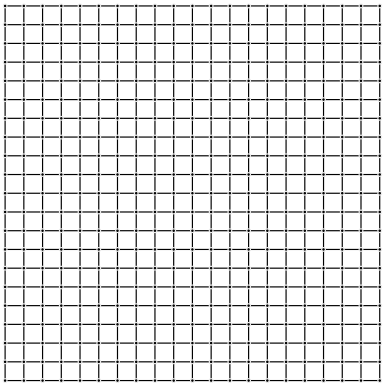
$$H_x(0, x, y) = 0$$

$$H_y(0, x, y) = -\sin(2\pi x)$$

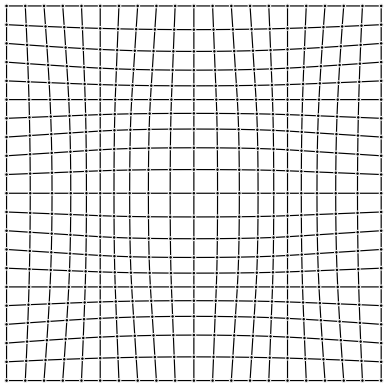
The exact solution is

$$E_z(t,x)(t,x,y) = \sin(2\pi(x-t))$$
$$H_x(0,x,y) = 0$$
$$H_y(0,x,y) = -\sin(2\pi(x-t))$$

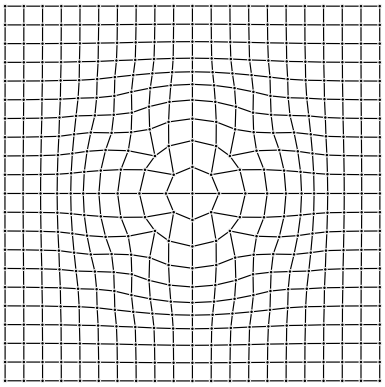
We solve the system with BFECC scheme based on least square θ-scheme from $t = 0$ to $t = 2.5$ with $\Delta t/\Delta x = 0.25$, and compare the solutions with exact solutions. The ratio $\Delta t/\Delta x$ is set to be 0.25 to make sure it greater than the CFL number for non-uniform grids. The problem is solved in four grids: (a) uniform rectangular grid, (b) non-rectangular grid obtained by smooth perturbation from (a), (c) non-rectangular grid with global circular grid deformation, and (d) non-rectangular grid with grid points shifted to a circular interface. The grids are shown in Figure-3.4 and the order of accuracy is shown in Table-3.3. We see the numerical order of accuracy are all above 2, verifying the effectiveness of the BFECC method on non-orthogonal grids.



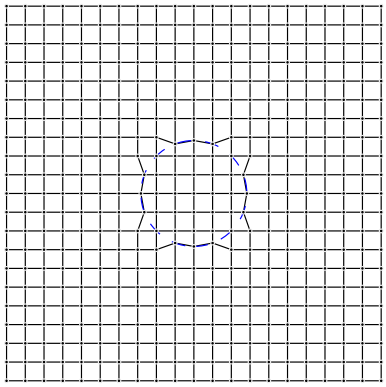
(a) Uniform rectangular grid.



(b) Non-rectangular, smoothly varying.



(c) Non-rectangular with global circular deformation.



(d) Non-rectangular with local deformation.

Figure 3.4: Grids: (a) Uniform rectangular; (b) (c) and (d) Non-orthogonal grids

Table 3.3: Order of accuracy for BFECC + least square θ-scheme at $T = 2.5$

Grid	(a)		(b)		(c)		(d)	
	Error	Order	Error	Order	Error	Order	Error	Order
20×20	5.843×10^{-2}	–	1.502×10^{-1}	–	6.429×10^{-2}	–	5.723×10^{-2}	–
40×40	8.16×10^{-3}	2.84	2.469×10^{-2}	2.61	1.070×10^{-2}	2.59	7.013×10^{-3}	3.03
80×80	1.269×10^{-3}	2.69	3.426×10^{-3}	2.85	2.413×10^{-3}	2.15	8.485×10^{-4}	3.05

3.10.4 2D wave absorption by perfectly match layers

This example demonstrate the effectiveness of our implementation of the perfectly matched layers. We solve the Maxwell's equations in the two dimensional $\mathrm{TM_z}$ mode in the whole $x - y$ plane. The initial condition is

$$E_z = \exp\left[-\frac{(x-0.5)^2 + (y-0.5)^2}{2 \times (0.1)^2}\right], \; H_x = H_y = 0$$

The solution for E_z is a radially symmetric wave propagating to infinity.

We simulate this problem with a computation domain $[0, 1] \times [0, 1]$ with perfectly matched layers surrounded. The figures below show the numerical solution using a 80×80 uniform rectangular grid, 40 layers of perfectly match layers on each side of the computation domain and artificial conductivity parameters $\sigma_x = \sigma_y = 800$. Figure-3.5 and Figure-3.6 shows the solution profile at $t = 0, 0.2, 0.4, ..., 2.2$, from which we see the wave propagates freely out of the computation domain with very small reflection. Figure-3.7 and Figure-3.8 show the total energy in the computation domain at different time, again verifying the effectiveness of the PML implementation. At $t = 1.0$ the total energy is reduced to 1% of the initial energy value, and at $t = 3.0$, it is reduced to 10^{-5} of the initial energy value.

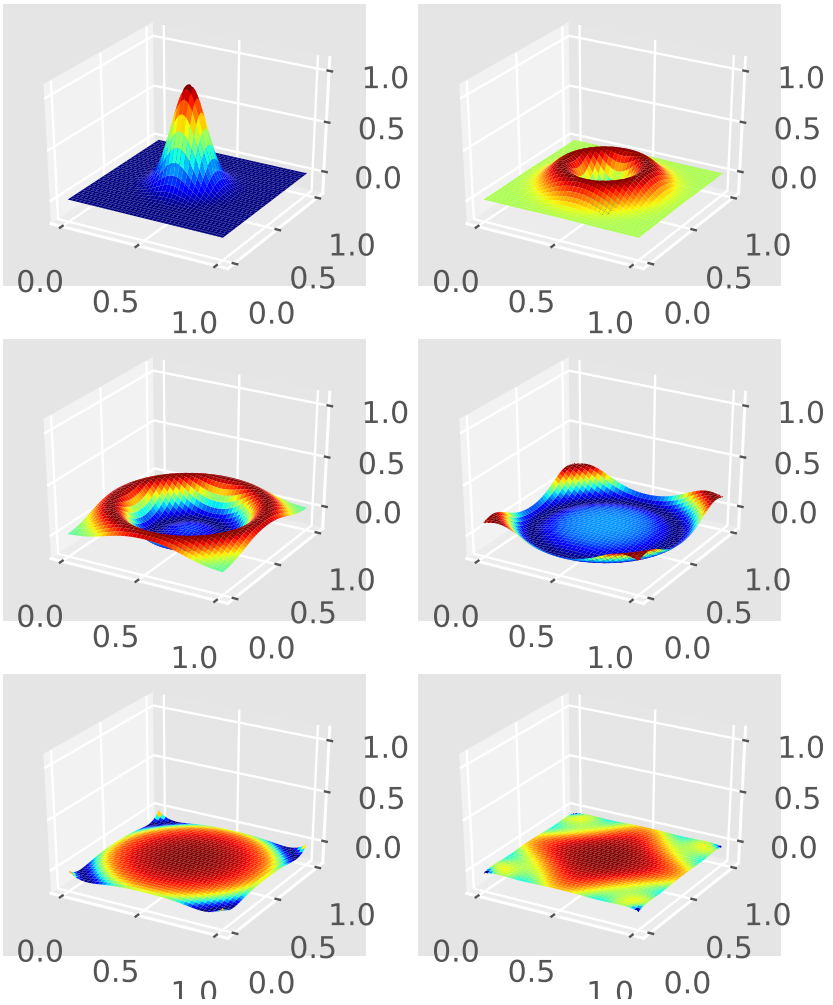


Figure 3.5: Solution profile at different time: from upper left to lower right, it shows E_z surfaces at $t = 0, 0.2, 0.4, 0.6, 0.8, 1.0$.

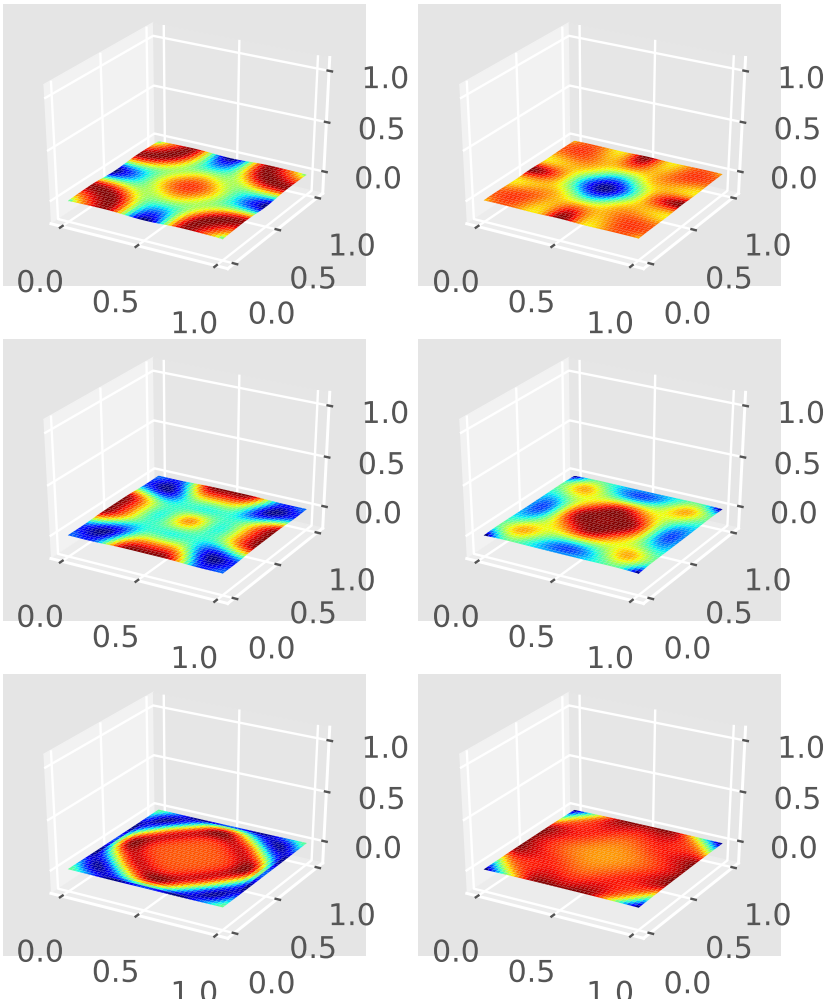


Figure 3.6: Solution profile at different time: from upper left to lower right, it shows E_z surfaces at $t = 1.2, 1.4, 1.6, 1.8, 2.0, 2.2$.

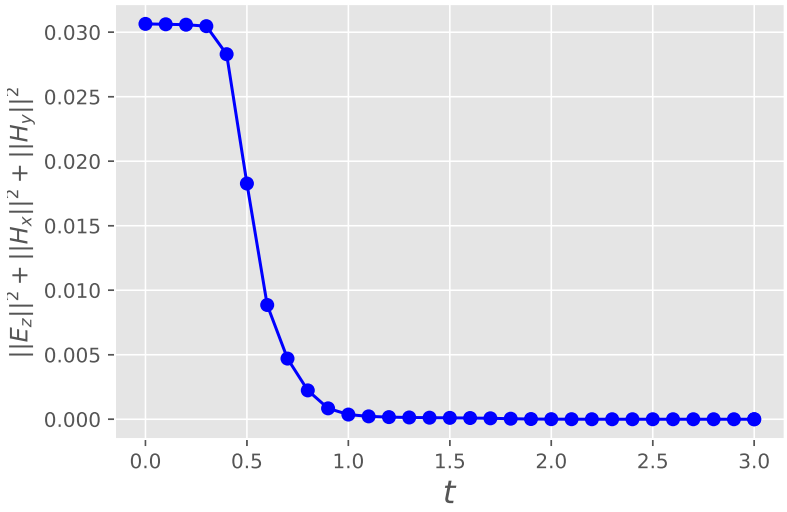


Figure 3.7: Total energy of the electromagnetic field versus time.

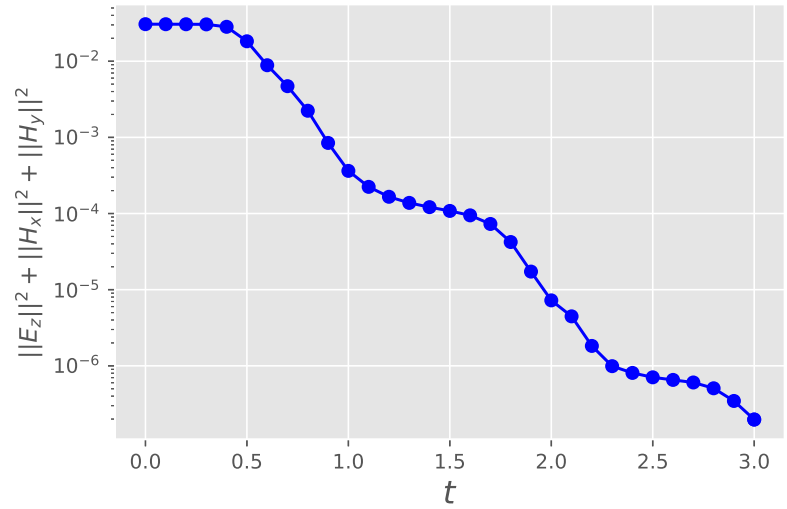


Figure 3.8: Total energy of the electromagnetic field versus time, semi-log plot: at $t = 1.0$, total energy is reduced to 1% of the initial value and $t = 3$, it is reduced to 10^{-5} of the initial value.

3.10.5 Scattering by a dielectric cylinder

In this example, we solve the 2D Maxwell's equations in $\mathrm{TM_z}$ mode with the BFECC + least square θ-scheme for the scattering problem by a dielectric cylinder.

$$\mu\frac{\partial H_x}{\partial t} = -\frac{\partial E_z}{\partial y}$$
$$\mu\frac{\partial H_y}{\partial t} = \frac{\partial E_z}{\partial x}$$
$$\epsilon\frac{\partial E_z}{\partial t} = \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}$$

The incident wave is a z-polarized plane wave travelling in the x direction, i.e. $(E_z)_{inc} = \sin(\omega(x-t))$, $(H_x)_{inc} = 0$ and $(H_y)_{inc} = -\sin(\omega(x-t))$, where $\omega = 2\pi/0.6$ is the angular frequency. The computation domain is $[0,1] \times [0,1]$. A dielectric cylinder with $\epsilon_1 = 2.25$ and $\mu_1 = 1$ and radius 0.24 is placed in the center of the computation domain. The surrounding medium has $\epsilon_0 = 1$ and $\mu_0 = 1$. Perfectly match layers are used as absorbing boundaries, and total-field/scattered-field formulation is used to introduce plane waves into the computation domain.

Two grids are used in computation: (a) a uniform rectangular grid is used and the material interface is approximated by stair-casing; and (b) a point shifted grid in which intersection points of the uniform rectangular grid and the material interface are computed and the closest rectangular grid points are moved to the intersection points, and it is shown in Figure-3.1 (a). We use a simple treatment for the material interface: if a grid point falls inside the dielectric cylinder, $\epsilon_1 = 2.25$ and $\mu_1 = 1$ are used during update of $\boldsymbol{E}$ and $\boldsymbol{H}$, otherwise, $\epsilon_0 = 1$ and $\mu_0 = 1$ are used. Better interface treatment is planned for our future work.

The BFECC + least square θ-scheme is used instead of the BFECC + least square central difference scheme is used. The larger numerical dissipation is helpful for the material

discontinuity. When The BFECC + least square central difference scheme is used, there are small oscillations presented in the numerical solution due to the material discontinuity.

Since the CFL condition for BFECC + least square θ-scheme only requires $\Delta t \leq \frac{\sqrt{3}}{\sqrt{(1/\Delta x)^2+(1/\Delta y)^2}}$, here we take $\Delta t = \Delta x = \Delta y$. Smaller Δt values have been experimented, giving similar results as presented here.

The numerical solution on the point-shifted grid at $t = 3.8$ is shown in Figure-3.9 and is compared with the analytic Mie solution [34] in Figure-3.10. The BFECC + least square θ-scheme scheme is able to generate smooth solutions without any stair casing oscillation. $t = 3.8$ is chosen since the solution seems to reach the steady state at this time. The scheme are applied for several thousands time steps (up to $t = 12$) and the solution remains stable.

The grid refinement analysis for numerical solutions on uniform rectangular grids and point shifted grids is shown in Table-3.4. Here the numerical solution on a 320×320 grid is taken as the approximated accurate solution, and all errors are computed with respect to this numerical solution. We see the BFECC scheme achieves second order accuracy.

Table 3.4: Order of accuracy for BFECC + least square θ-scheme at $T = 3.8$

Grid	uniform rectangular		non-rectangular (d)	
	Error	Order	Error	Order
20×20	0.2741	–	0.4208	–
40×40	0.0789	1.80	0.1301	1.69
80×80	0.0148	2.41	0.0341	1.93
160×160	0.0037	2.00	0.0068	2.33

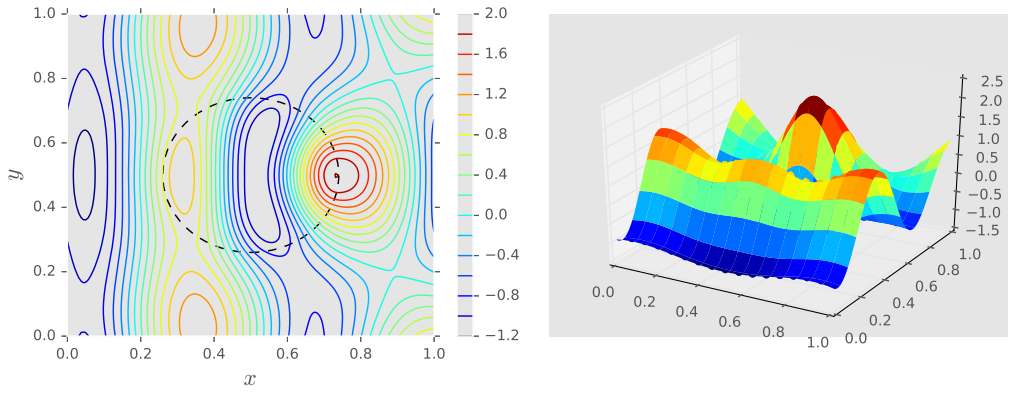


Figure 3.9: BFECC + least square θ-scheme solution at $t = 3.8$. Left: contour plot of E_z; Right: surface plot of E_z

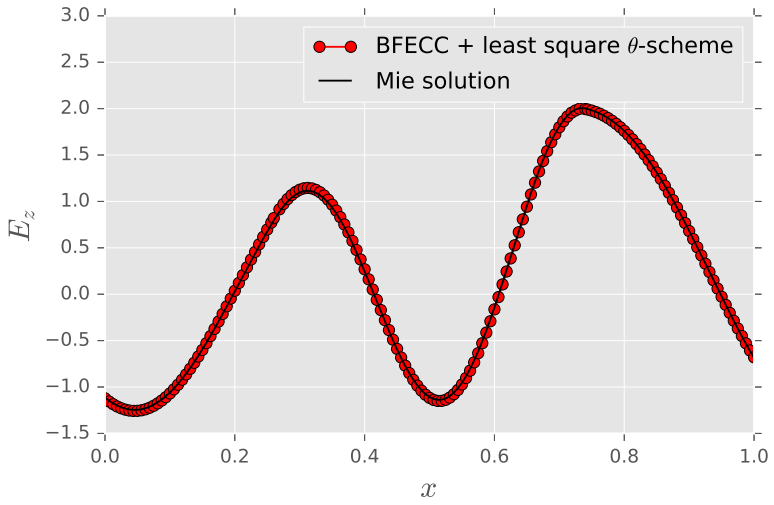


Figure 3.10: Slice of E_z with $y = 0.5$ at $t = 3.8$, compared with the analytic Mie solution.

3.10.6 Scattering by a dielectric object of complicated shape

In this example, the BFECC + least square θ-scheme is applied to solve a scattering problem by a dielectric object of more complicated shape. The grid and material setup is the same as the above example, except the object has a more complicate shape with sharp corners and cavities inside, as shown in Figure-3.11. The two grids used for computations are (a) a uniform rectangular grid with staircasing approximation for material interface and (b) a

point shifted grid, see Figure-3.1 (b).

The object and the contour plot of E_z at $t = 3.6$ is shown in Figure-3.11. Taking the 320×320 numerical solution as reference, the numerical errors are shown in Table-3.5. Again we see the BFECC scheme is stable and has second order accuracy.

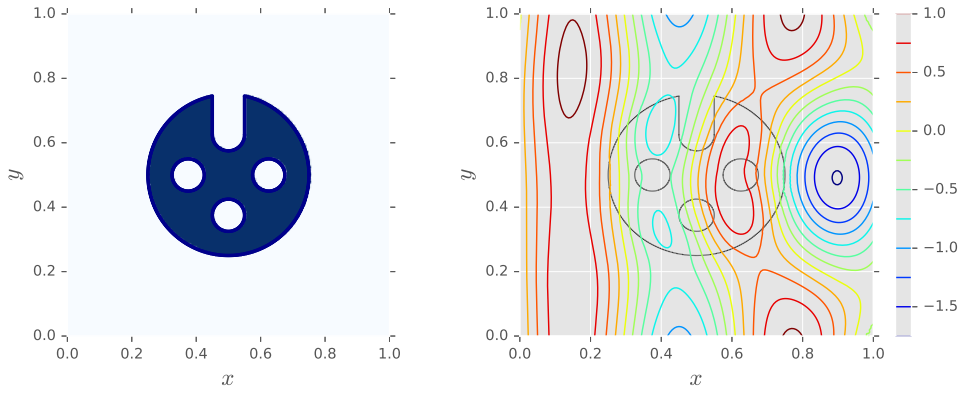


Figure 3.11: Scattering by a object of complicated shape. Left: shape of the object; Right: contour plot of E_z at $t = 3.6$.

Table 3.5: Grid refinement analysis for BFECC + least square θ-scheme at $T = 3.6$

Grid	uniform rectangular		non-rectangular (d)	
	Error	Order	Error	Order
20×20	4.200×10^{-1}	–	4.384×10^{-1}	–
40×40	1.159×10^{-1}	1.86	1.160×10^{-1}	1.92
80×80	3.618×10^{-2}	1.68	3.700×10^{-2}	1.65
160×160	8.116×10^{-3}	2.16	8.830×10^{-3}	2.07

In summary, this chapter focuses on the application of the BFECC method to the Maxwell's equations. On uniform orthogonal grids, the BFECC + central difference or BFECC + Lax-Friedrichs schemes are provably second order accurate and have larger CFL number than the classical Yee scheme. On non-orthogonal or unstructured grids, the BFECC method is applied to a first order scheme based on least square gradient ap-

proximation. Numerical examples demonstrate the effectiveness of the BFECC schemes for Maxwell's equations. In particular the BFECC + least square central difference scheme and BFECC + least square θ-scheme are easy to implement on non-orthogonal grids and have second order accuracy in the numerical examples we tested. We plan to test the BFECC schemes on non-orthogonal and unstructured grids for more application problems in future.

CHAPTER 4
BFECC METHOD FOR SCALAR CONSERVATION LAWS AND A CONSERVATIVE BFECC LIMITER

In this section, we discuss our effort in applying BFECC method to hyperbolic PDEs that arises from conservation laws. Two distinct features of conservation laws is the presence of discontinuities and the importance of preserving conserved quantities. We focus on designing BFECC schemes that can correctly handle discontinuities and preserve conserved quantities.

For high order (second order or higher) numerical schemes of conservation laws, the presence of discontinuity often results in spurious oscillations near the discontinuities. In classical schemes such as MUSCL [8], ENO and WENO [9, 10], the spurious oscillations are avoided using limiters or specially designed interpolation method. These techniques usually requires complicated function reconstruction or high order interpolation. The numerical oscillation is observed in BFECC schemes for advection equations and a limiter is proposed to reduce the spurious oscillation [35]. The limiter is based on the back-and-forth error correction idea, and it doesn't require complicated function reconstruction, making the limiter very efficient to use. The limiter doesn't preserve conserved quantities, though. For numerical solution of conservation laws with discontinuities, this non-conservation could pose a serious problem. For example, near a discontinuity, the numerical solution loses its order of accuracy, and non-conservation means quantity can be lost or created significantly at the discontinuity, making the solution to lose accuracy in the whole domain. In the first two sections of this chapter, we modify the limiter in [35] to make it conservative. The limiter is applied to advection equations to show it effectiveness.

To solve other equations from conservation laws, we extend the BFECC method and the conservative limiter to finite volume schemes. The BFECC + limiter scheme is then

use to solve the inviscid Burgers' equation, and to solve the convection part in the two dimensional Vlasov-Poisson system, viscous Burgers' equation and the Korteweg-de Vries (KdV) equation.

4.1 Limiting by truncation

Consider advection equation

$$\frac{\partial u}{\partial t} + \boldsymbol{v} \cdot \nabla u = 0 \tag{4.1}$$

where $\boldsymbol{v}(x,t)$ is a known vector field.

The BFECC scheme could introduce spurious oscillation near the discontinuities of the numerical solution, see Figure-4.2 for an example. The following nonlinear limiter was introduced to eliminate these oscillations [35]. The limiter is based on comparing relative size of error terms.

In this section, we will use the Courant-Issacson-Rees (CIR) [2] scheme as the underlying scheme for the advection equation 4.1. CIR schemes is a first order unconditionally stable scheme (both in l^2 and l^∞ sense), and it updates numerical solutions by tracing back along the characteristics, finding the intersection of the characteristic with $t = t^n$, and interpolating from neighboring U^n values.

Let $e^{(1)} = \frac{1}{2}(I - L^*L)U^n$ be the backward error compensation term. Define

$$e^{(2)} = (I - L^*L)e^{(1)}$$

Then we have the following relation between the relative magnitudes of $e^{(1)}$ and $e^{(2)}$:

Proposition 1. *Suppose L is a linear scheme for equation (4.1), L^* is the same scheme for the time reversed equation, $\rho_{L^*} = \overline{\rho_L}$ and $|\rho_L| \leq \sqrt{2}$, then $||e^{(2)}||_2 \leq ||e^{(1)}||_2$.*

Proof. By the assumption,

$$\hat{e}^{(2)} = (1 - |\rho_L|^2)\hat{e}^{(1)}$$

□

We can use the relative comparison of $|e_i^{(2)}|$ and $|e_i^{(1)}|$ as a detector for discontinuity. In particular, when BFECC is applied the CIR scheme, we have:

$$\left(L^* L e^{(1)}\right)_i = \sum_{j \in I} c_i e_{i+j}^{(1)}$$

Here the set $i + I$ is consists of indices that is involved in computation for $\left(L^* L e^{(1)}\right)_i$, and typically it includes indices of grid points x_j such that $|x_j - x_i| \leq K|v|\Delta t$, where K is a fixed positive integer.

Moreover, we have the following theorem [35]

Theorem 8 ((Theorem 3.3 in [35])). *Suppose the linear scheme $\mathcal{L}$ is consistent, monotone and at least of first order accuracy. If $|e_i^{(1)}|$ is a a local maximum, and $(\mathcal{L}^* \mathcal{L} e^{(1)})_i$ has the same sign as $e_i^{(1)}$, then $|e_i^{(2)}| \leq |e_i^{(1)}|$.*

From the theorem we see $|e_i^{(2)}| > |e_i^{(1)}|$ means there is a $j^* \in I$ such that $|e_{i+j^*}^{(1)}|$ is significantly larger than $|e_{i+j}^{(1)}|$ for $j \in I$ $j \neq j^*$. Furthermore, this large $|e_{i+j^*}^{(1)}|$ can only be caused by a rapid change in the value of U in the vicinity of i. Therefore, $|e_i^{(2)}| > |e_i^{(1)}|$ indicates the presence of a discontinuity in the vicinity of i. An example is presented in Figure-4.1, in which we see the position where $|e_i^{(2)}| > |e_i^{(1)}|$ corresponds to the position of overshooting/undershooting of the numerical solution.

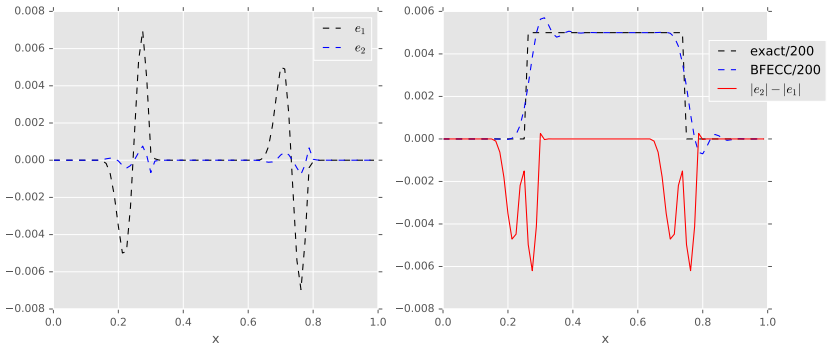


Figure 4.1: $|e_i^{(2)}| > |e_i^{(1)}|$ indicates overshooting/undershooting

Based on this observation, the following limiting algorithm is proposed in [35].

1 Detect overshooting/undershooting: Recall $e^{(1)} = \frac{1}{2}(I - \mathcal{L}^*\mathcal{L})U^n$, define $e^{(2)} = (I - \mathcal{L}^*\mathcal{L})e^{(1)}$. If $|e_i^{(2)}| > |e_i^{(1)}|$ at some grid point, then overshooting/undershooting is likely to occur at an adjacent grid point of i.

2 Truncate $e^{(1)}$: make a copy $\tilde{e}^{(1)}$ of $e^{(1)}$, at every grid point x_i such that $|e_i^{(2)}| > |e_i^{(1)}|$, truncate the error so that $\tilde{e}_j^{(1)} = \text{minmod}\left(e_i^{(1)}, \tilde{e}_j^{(1)}\right)$ for x_j adjacent to x_i.

3 Correct the solution using using the truncated $e^{(1)}$ and compute the solution at time level $n+1$, $U^{n+1} = L(U^n + e^{(1)})$.

This truncation limiter reduces oscillations near discontinuities effectively, but it doesn't guarantee to preserve the conserved quantity of the equation even if the underlying scheme for BFECC is conservative. We explore a modification of the limiter in the next section to make it conservative.

4.2 Conservative limiting

BFECC + CIR is a conservative scheme when the velocity field is constant. Limiting by truncation destroys the conservation property. We can restore the conservation property easily by redistributing the mass loss.

- Step 1. Compute the back-and-forth error $e^{(1)} = \frac{1}{2}\left(I - \mathcal{L}^*\mathcal{L}\right)U^n$, $e^{(2)} = (I - \mathcal{L}^*\mathcal{L})e^{(1)}$, truncate $e^{(1)}$ as the third step of truncation limiter to get $\tilde{e}^{(1)}$.

- Step 2. Let $\Delta e^{(1)} = e^{(1)} - \tilde{e}^{(1)}$. At every grid point x_i such that $\Delta e_i^{(1)} \neq 0$, redistribute $\Delta e_i^{(1)}$ to its neighbor grid points: .

  Proper redistribution: suppose $\Delta e_i^{(1)} \neq 0$, redistribute $\Delta e_i^{(1)}$ to its neighbor grid points, by solving

$$\begin{aligned} &\min \max\{|\tilde{e}_{i-1}^{(1)} + x|, |\tilde{e}_i^{(1)} + y|, |\tilde{e}_{i+1}^{(1)} + z|\} \\ \text{s.t. } & x + y + z = \Delta e_i^{(1)} \end{aligned}$$

  For more efficient redistribution, we replace the l^∞ optimization by l^2 optimization:

$$\begin{aligned} &\min \left(\tilde{e}_{i-1}^{(1)} + x\right)^2 + \left(\tilde{e}_i^{(1)} + y\right)^2 + \left(\tilde{e}_{i+1}^{(1)} + z\right)^2 \\ \text{s.t. } & x + y + z = \Delta e_i^{(1)} \end{aligned}$$

  Denote the error compensation term after redistribution $e^{(c)}$.

- Step 3. Set $e^{(1)} = e^{(c)}$.

- Step 4. Repeat step 1 - 3 until there is no grid point x_i at which $|e_i^{(2)}| > |e_i^{(1)}|$. Use the final $\tilde{e}^{(1)}$ as the error compensation term in the last step of BFECC.

The redistribution step is designed to make $|\tilde{e}^{(1)}|$ as "smooth" as possible. By Theorem-8, this reduces the number of indices i for which $|e_i^{(2)}| > |e_i^{(1)}|$. Smoother $|e^{(1)}|$ also makes the corrected solution $U^{n+1} = \mathcal{L}\left(U^n + e^{(1)}\right)$ less oscillatory.

An example of the conservative limiter is shown in Figure-4.2. Here $u_t + u_x = 0$ is solved on $[0, 1]$ with the CIR, BFECC + CIR and BFECC + CIR + conservative limiter with periodic boundary condition. The initial data is $u(0, x) = 1.0$ if $0.25 \leq x \leq 0.75$.

Solution is shown at $T = 1.0$. $\Delta t/\Delta x = 0.5$. The conservative limiter successfully reduces the spurious oscillations. The same equation is solved with a smooth initial data $u(0, x) = 1 + 0.5\sin(2\pi x)$ to test the order of accuracy. As shown in Table-4.1, the limiter doesn't change the order of accuracy of the scheme.

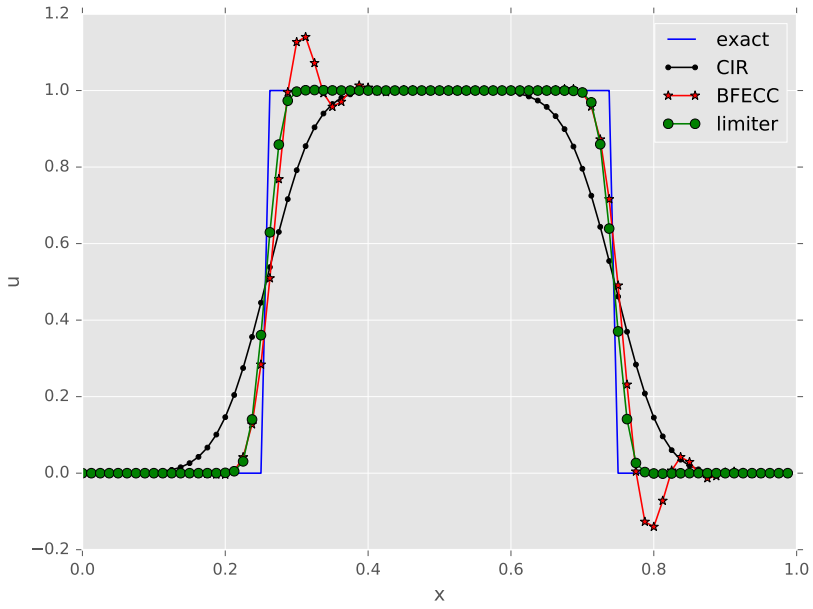


Figure 4.2: BFECC with the conservative limiter.

Table 4.1: Grid refinement analysis. Numerical solutions at $T = 20$, $\Delta t/\Delta x = 2.2$.

	CIR		BFECC		Limiter	
Mesh	Error	Order	Error	Order	Error	Order
40	2.85×10^{-2}	–	1.17×10^{-3}	–	1.26×10^{-3}	–
80	1.18×10^{-2}	1.27	2.23×10^{-4}	2.50	2.23×10^{-4}	2.50
120	6.86×10^{-3}	1.35	8.09×10^{-5}	2.50	8.09×10^{-5}	2.50
160	4.58×10^{-3}	1.41	3.94×10^{-5}	2.50	4.03×10^{-5}	2.42
200	3.33×10^{-3}	1.43	2.56×10^{-5}	2.50	2.25×10^{-5}	2.60

4.3 A conservative BFECC solver for the Vlasov-Poisson equation

In this section, we apply the conservative BFECC limiter in Section-4.2 for the Vlasov-Poisson equation. The goal here is to use the BFECC solver as an efficient and easy-to-implement alternative to other solvers for the Vlasov-Poisson equation.

Under proper assumptions(collisionless, nonrelativisitic, and no magnetic field), the evolution of particle density function $f(x, v, t)$ for a dilute plasma can be described by the Vlasov-Poisson equation

$$\partial_t f + v \cdot \nabla_x f + \frac{-eE}{m} \cdot \nabla_v f = 0$$

Here $f(x, v, t)$ is the particle number density at time t, position x and with velocity v, e is the charge of the underlying particle (electron) and E is the electric field. E is coupled to the particle density function through the Poisson equation

$$\begin{aligned} E &= -\nabla\phi \\ \Delta\phi &= -\rho \\ \rho(x,t) &= \rho_i(x,t) - e\int_{\mathbb{R}^3} f(x,v,t)dv \end{aligned}$$

where $\rho_i(x, t)$ is the density of charges from ions.

In the following, we assume there is a uniform density of charges from ions and rewrite the system of equations in normalized units.

$$\begin{aligned} &\partial_t f + v \cdot \nabla_x f + E \cdot \nabla_v f = 0 \\ &E = -\nabla\phi \\ &\Delta\phi = -\rho \\ &\rho = \int_{\mathbb{R}^3} f(x,v,t)dv - 1 \end{aligned}$$

Our solver is based on Strang splitting of the Vlasov-Poisson equation, to update numerical solution at time t_n to time $t_{n+1} = t_n + \Delta t$ we do:

- Step 1. Update $\partial_t f + v \cdot \nabla_x f = 0$ for $\Delta t/2$ using the BFECC solver with conservative limiter with initial condition f^n, get the updated density function $f^{n+1/2}$.
- Step 2. Update the electric field $E^{n+1/2}$ by solving the Poisson equation with updated density $f^{n+1/2}$; then solve $\partial_t f + E^{n+1/2} \cdot \nabla_v f = 0$ for Δt using the BFECC solver with conservative limiter with initial condition $f^{n+1/2}$, get updated density function $\tilde{f}^{n+1/2}$.
- Step 3. Update $\partial_t f + v \cdot \nabla_x f = 0$ for $\Delta t/2$ using the BFECC solver with conservative limiter with initial condition $\tilde{f}^{n+1/2}$, get the updated density function f^{n+1}.

Proposition 2. *The above numerical scheme is a second order (in space and time) scheme for the Vlasov-Poisson equation that preserves total particle number.*

Proof. In each step of the scheme, when f is updated by the BFECC solver with conservative limiter, the velocity field is constant. So by the property of BFECC solver with conservative limiter, it conserves total particle number. Therefore, the total particle number is conserved.

For the order of accuracy: since at each step the velocity field is constant, BFECC solver with conservative limiter is second order accurate in space and time. Strang splitting introduces addition error of third order, so the total order of accuracy is still second order.

□

In the following, we demonstrate a few numerical examples.

Example 1. *(1D Weak Landau Damping)*

In this numerical example, we study the solution that corresponds to the weak Landau damping. It has analytical theory and is often used as a benchmark problem for numerical

schemes for the Vlasov-Poisson system [36]. We consider the initial condition:

$$f_0(x,v) = \frac{1}{\sqrt{2\pi}}(1+\alpha\cos(kx))e^{-v^2/2}$$

This is a small perturbation away from the equilibrium solution.

We first run the case $\alpha = 0.001$, $k = 0.4$ *and compare it with an approximate analytic solution. The results agrees well (up to a multiplicative factor due to using different normalization).*

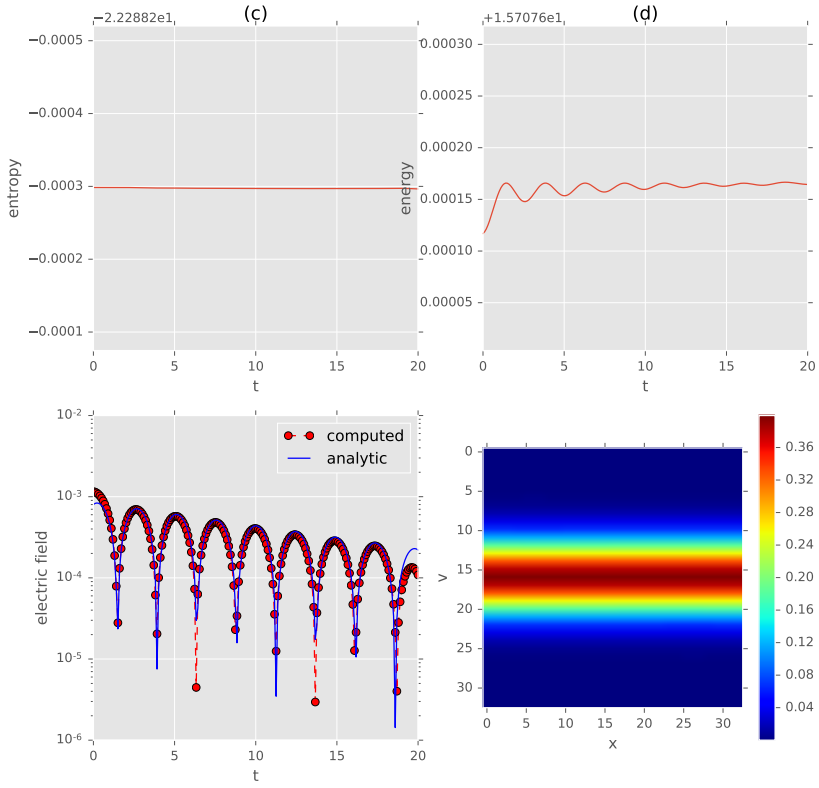


Figure 4.3: Weak Landau Damping: entropy, energy, electric field and density function

Example 2. *(1D two stream instability [37]) The initial condition is:*

$$f_0(x,v) = \frac{2}{7\sqrt{2\pi}}(1+5v^2)\left[1+\alpha\left(\frac{\cos(2kx)+\cos(3kx)}{1.2}+\cos(kx)\right)\right]e^{-v^2/2}$$

with $\alpha = 0.01$, $k = 0.5$, $L_x = 2\pi/k = 4\pi$.

In the following, we use $N_x = 32$, $N_v = 32$, *instead of the higher values in literature in order to make the computation faster. We use* $v_{\max} = 5$ *and* $\Delta t = 1/8$, *the same as [38].*

In the following, we plotted the l_1 *and* l_2 *norm of the* f, *entropy and total energy. We observe that our scheme preserves the* l_1 *norm, the error in* l_2 *norm of* f *and* $entropy$ *is comparable to higher order schemes in [38] (3rd and 5th order), and the error in total energy is also very small.*

The plot for electric field and phase space plot of f *also agree very well with results from high order schemes in [38]. In particular, for the phase space plot of* f, *we also get details resolved at the center of mixing.*

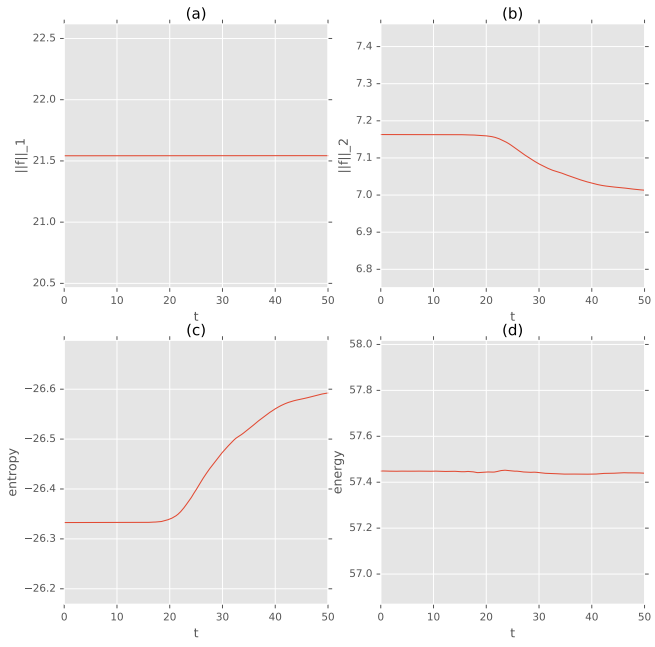


Figure 4.4: Two stream instability: norm, entropy and energy

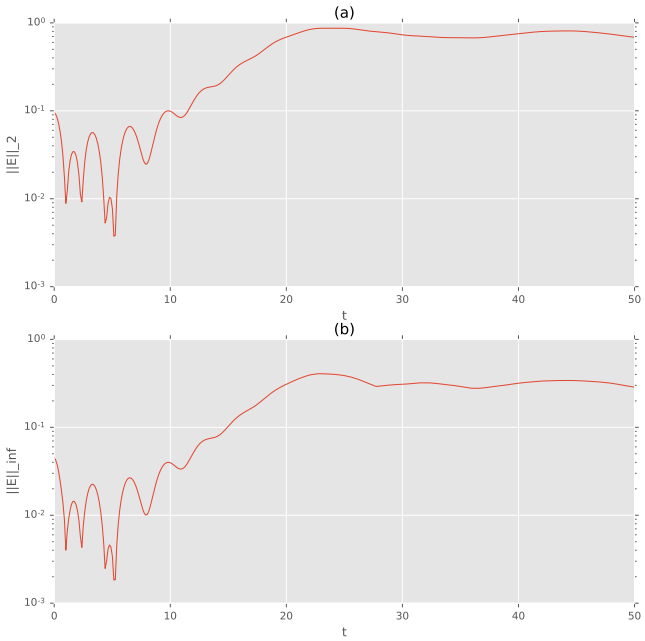


Figure 4.5: Two stream instability: electric field

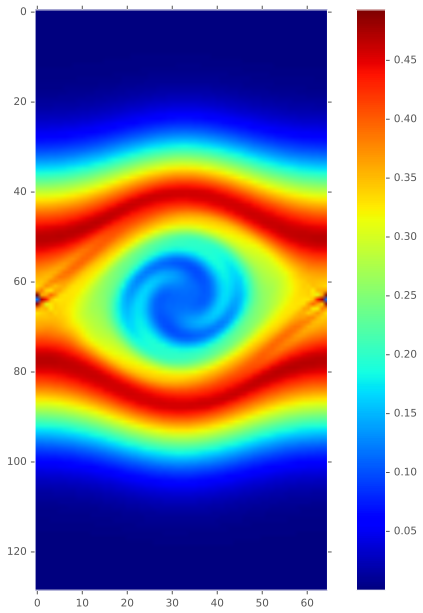


Figure 4.6: Two stream instability: density function at T = 50

4.4 BFECC for semi-Lagrangian finite volume scheme

In this section, we discuss a simple generalization of BFECC method to the semi-Lagrangian finite volume scheme. In [38], a semi-Lagrangian finite volume has been used for the Vlasov-Poisson system to ensure conservation and achieve high spatial order of accuracy at the same time. We show in this section we can apply combine BFECC and the semi-Lagrangian finite volume scheme to further improve the accuracy.

Consider the one dimensional advection equation

$$u_t + au_x = 0$$

with periodic boundary condition on $[0, L]$.

In a finite volume scheme, the cell average

$$\bar{u}_i^n = \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u(x, t_n) dx$$

is updated at each step.

At time $t = t_n$, the cell averages $\{\bar{u}_i^n\}_{i=1,2,\ldots,m}$ are known. Now, we solve for the cell average at time $t = t_{n+1}$:

$$\begin{aligned}
\bar{u}_i^{n+1} &= \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u(x, t_{n+1}) dx \\
&= \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} u(X(t_n; x, t_{n+1}), t_n) dx \\
&= \frac{1}{\Delta x} \int_{y_{i-\frac{1}{2}}}^{y_{i+\frac{1}{2}}} u(\xi, t_n) J d\xi
\end{aligned}$$

where $x = X(t; z, t_n + 1)$ is the characteristic line and $J = \left(\frac{dX(t_n;x,t_{n+1})}{dx}\right)^{-1}$ is the Jacobian, and $y_{i\pm1/2} = X(t_n; x_{i\pm1/2}, t_n + 1)$. When a is a constant, we have $X(t_n; x, t_{n+1}) = x - a\Delta t$, $J = 1$ and $y_{i\pm1/2} = x_{i\pm1/2} - a\Delta t$. So

$$\bar{u}_i^{n+1} = \frac{1}{\Delta x} \int_{y_{i-\frac{1}{2}}}^{y_{i+\frac{1}{2}}} u(x, t_n) dx$$

The right hand side is approximated using an interpolation: define the primitive function $U(x) = \int_0^x u(\xi, t_n) d\xi$, and define

$$U_i^n = U^n(x_{i+1/2}) = \Delta x \sum_{j=1}^{i} \bar{u}_j^n.$$

Then

$$\frac{1}{\Delta x} \int_{y_{i-\frac{1}{2}}}^{y_{i+\frac{1}{2}}} u(x, t_n) dx = \frac{1}{\Delta x} \left(U^n(y_{i+1/2}) - U^n(y_{i-1/2}) \right) \tag{4.2}$$

And values of $U^n(y_{i\pm1/2})$ are approximated by interpolated value from $\{U_i^n\}$. This scheme is clearly conservative from equation-4.2.

In the semi-Lagrangian finite volume scheme, the quantity that is updated at each step is cell average $\bar{u}_i^n$ instead of point values u_i^n. Replacing u_i^n by $\bar{u}_i^n$ in the BFECC method, we obtain the BFECC semi-Lagrangian finite volume scheme.

In the following, we show that when fixed stencil interpolation is used for the interpolation step in semi-Lagrangian finite volume scheme, we can directly apply Theorem-3 in Chapter 2 to show that odd order scheme can be improved in order of accuracy by one.

Now, we compute the Fourier symbol. Suppose $\bar{u}_j^n = c_k^n e^{2\pi i k x_j}$. Then:

$$U_i^n = U^n(x_{i+1/2}) = \Delta x \sum_{j=1}^{i} \bar{u}_j^n = \Delta x c_k^n \sum_{j=1}^{i} e^{2\pi i k x_j}$$

With a fixed stencil interpolation, the value of $U^n(y_{i+1/2})$ is a linear combination of surrounding U_i^n values.

$$U^n(y_{i+1/2}) = \sum_{j\in J} \tilde{c}_j U_{i+j}^n$$
$$U^n(y_{i-1/2}) = \sum_{j\in J} \tilde{c}_j U_{i+j-1}^n$$

Therefore

$$\frac{1}{\Delta x}\left(U^n(y_{i+1/2}) - U^n(y_{i-1/2})\right) = \sum_{j\in J} \tilde{c}_j \left(c_k^n \sum_{l=1}^{i+j} e^{2\pi i k x_l} - c_k^n \sum_{l=1}^{i+j-1} e^{2\pi i k x_l} \right) = c_k^n \sum_{j\in J} \tilde{c}_j e^{2\pi i k x_{i+j}}$$

Therefore the Fourier symbol is

$$\rho_L(k) = \sum_{j\in J} \tilde{c}_j e^{2\pi i k j \Delta x}$$

For the time-reversed scheme L^*, at the interpolation step, the offsets are in the other

direction, therefore,

$$U^n(y_{i+1/2}) = \sum_{j\in J} \tilde{c}_j U^n_{i-j}$$
$$U^n(y_{i-1/2}) = \sum_{j\in J} \tilde{c}_j U^n_{i-j-1}$$

And a similar calculation shows that

$$\rho_{L^*}(k) = \sum_{j\in J} \tilde{c}_j e^{-2\pi i k j \Delta x} = \bar{\rho}_L(k)$$

Apply Theorem-3 in Chapter 2, we see BFECC improves the accuracy of an odd order scheme by one order.

4.5 BFECC for inviscid Burgers' equation

In this section, we present the numerical results of the BFECC method applied to a nonlinear conservation law equation. We consider the inviscid Burgers equation in one dimension

$$u_t + uu_x = 0 \tag{4.3}$$

it can be written in the conservation law form

$$u_t + (f(u))_x = 0 \tag{4.4}$$

where $f(u) = \frac{1}{2}u^2$.

We consider the finite volume scheme for the conservation law, define

$$\bar{u}^n_i = \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} u(x, t^n)dx$$

Integrate the equation over $[x_{i-1/2}, x_{i+1/2}] \times [t^n, t^{n+1}]$, we get:

$$\int_{x_{i-1/2}}^{x_{i+1/2}} u(x, t^{n+1})dx - \int_{x_{i-1/2}}^{x_{i+1/2}} u(x, t^n)dx = -\left(\int_{t^n}^{t^{n+1}} f(u(x_{i+1/2}, t))dt - \int_{t^n}^{t^{n+1}} f(u(x_{i-1/2}, t))dt \right)$$

Therefore

$$\bar{u}_i^{n+1} = \bar{u}_i^n - \frac{1}{\Delta x}\left(\int_{t^n}^{t^{n+1}} f(u(x_{i+1/2}, t))dt - \int_{t^n}^{t^{n+1}} f(u(x_{i-1/2}, t))dt \right)$$

In Godunov scheme, we approximate the solution by piecewise constant function that equals to $\bar{u}_i^n$ in cell $[x_{i-1/2}, x_{i+1/2}]$, then at the boundary of two cells, we can solve a Riemann problem to get $u(x_{i-1/2}, t)$ for $t \in [t^n, t^{n+1}]$. With this piecewise constant approximation, the flux term $\int_{t^n}^{t^{n+1}} f(u(x_{i+1/2}, t))dt$ turns out to be a function of $\bar{u}_i^n$ and $\bar{u}_{i+1}^n$, which is referred to as the numerical flux function $\hat{f}$. Let U_j^n be the numerical approximation of $\bar{u}_j^n$, the Godunov scheme can be written as:

$$\bar{U}_i^{n+1} = \bar{U}_i^n - \frac{\Delta t}{\Delta x}\left(\hat{f}(\bar{U}_i^n, \bar{U}_{i+1}^n) - \hat{f}(\bar{U}_{i-1}^n, \bar{U}_i^n) \right) \tag{4.5}$$

Instead of using the numerical flux function obtained from solving a Riemann problem, we can also use other suitable numerical flux function. In the following, we use the Lax-Friedrichs flux function:

$$\hat{f}(u_j, u_{j+1}) = \frac{1}{2}\left(f(u_j) + f(u_{j+1}) - \alpha(u_{j+1} - u_j) \right) \tag{4.6}$$

where $\alpha = \max_u |f'(u)|$.

Combining scheme (4.5) with the Lax-Friedrich flux function (4.6), we obtain a first order conservative scheme $\mathcal{L}$ for the one dimensional conservation law equation. Note when $f(u)$ is a linear function, i.e. $f(u) = au$ for some constant a, this scheme is simply the upwind scheme for equation $u_t + au_x = 0$.

Applying the BFECC method to this scheme is similar to the above case of semi-Lagrangian finite volume scheme:

1 Apply $\mathcal{L}$ to equation $u_t + (f(u))_x = 0$ to update $\{U_j^n\}$ to $\{\tilde{U}_j^{n+1}\}$;

2 Apply $\mathcal{L}$ to the time reversed equation $u_t - (f(u))_x = 0$ to update $\{\tilde{U}_j^{n+1}\}$ to $\{\tilde{U}_j^n\}$, and calculate the error compensation term $e^{(1)} = \frac{1}{2}\left(U^n - \tilde{U}^n\right)$;

3 Apply $\mathcal{L}$ to equation $u_t + (f(u))_x = 0$ to update $\{U_j^n + e_j^{(1)}\}$ to $\{U_j^{n+1}\}$.

We will refer this scheme as $\mathcal{L}_{BFECC}$.

When discontinuities are presented in the solution, the BFECC scheme could produce spurious oscillation. We can apply the previous conservative limiting method to the BFECC scheme to reduce such spurious oscillation.

For a linear equation $u_t + au_x = 0$, by Theorem-3 in Chapter 2, $\mathcal{L}_{BFECC}$ is second order accurate. It is conservative since the Godunov scheme is. For a nonlinear equation $u_t + (f(u))_x = 0$, the order of accuracy analysis is currently not available. Instead, in the following, we demonstrate the accuracy improvement through a numerical example of the inviscid Burgers equation.

Example 3. *Consider the inviscid Burgers equation*

$$u_t + uu_x = 0, t > 0, x \in [0, 1]$$

with initial condition

$$u(x, 0) = 1 + 0.1\sin(2\pi x)$$

and periodic boundary condition.

We use Godunov scheme with Lax-Friedrich flux $\mathcal{L}$, BFECC scheme $\mathcal{L}_{BFECC}$ and BFECC scheme with conservative limiter to numerically solve this equation. In order to compare accuracy, we also use the second order MUSCL [8] scheme to solve the equation.

Exact solutions can be found by propagating the initial date along the characteristic. From the exact solution, we know no shock is developed until $t = 2.5$. We solve the equation upto $T = 1.0$, and check the order of accuracy by comparing the numerical solutions with the exact solution. First compute the numerical solution with MUSCL scheme on a very fine mesh ($\Delta x = 2^{-11}$ and $\Delta t = 2^{-13}$) and use it as the high accuracy approximation for the exact solution. Then we solve the equation on meshes of different resolution and compute the l^2 error at each mesh size. The result is summarized in the table 4.2. We see BFECC and BFECC with conservative limiter are second order schemes

Table 4.2: Order of accuracy for Burgers equation, $\Delta t/\Delta x = 0.25$

	Godunov		BFECC		Limiter	
Mesh	Error	Order	Error	Order	Error	Order
64	2.37×10^{-3}	–	1.14×10^{-4}	–	1.04×10^{-4}	–
128	9.62×10^{-4}	1.30	1.96×10^{-5}	2.54	1.70×10^{-5}	2.62
256	3.73×10^{-4}	1.37	3.45×10^{-6}	2.51	3.28×10^{-6}	2.36
512	1.39×10^{-4}	1.42	7.75×10^{-7}	2.15	7.72×10^{-7}	2.09

Example 4. *Consider the inviscid Burgers equation*

$$u_t + uu_x = 0, t > 0, x \in [0, 1]$$

with initial condition

$$u(x, 0) = 1 + \sin(2\pi x)$$

and periodic boundary condition.

A shock develops at $t = 0.25$. We solve the equation upto $T = 1.0$ to see how the scheme tracks the shock. We use Godunov scheme with Lax-Friedrich flux $\mathcal{L}$, BFECC scheme $\mathcal{L}_{BFECC}$ and BFECC scheme with conservative limiter to numerically solve this equation. For comparison, we solve the equation with the MUSCL scheme. In addition, we

also show the numerical solution obtained using BFECC with a nonconservative limiter. The results are shown in figure 4 and figure 4.

The BFECC scheme gives a sharper shock profile than the Godunov scheme, but introduces some overshooting and undershooting near the shock. The BFECC with conservative limiter removes the overshooting/undershooting while keeping the sharp shock profile, and its results is very close to the second order MUSCL scheme. In fact, its shock profile seems to be even sharper than the MUSCL scheme. Finally, conservation is important for having the right speed of the shock, as shown in the comparison of nonconservative limiter, conservative limiter and the MUSCL scheme.

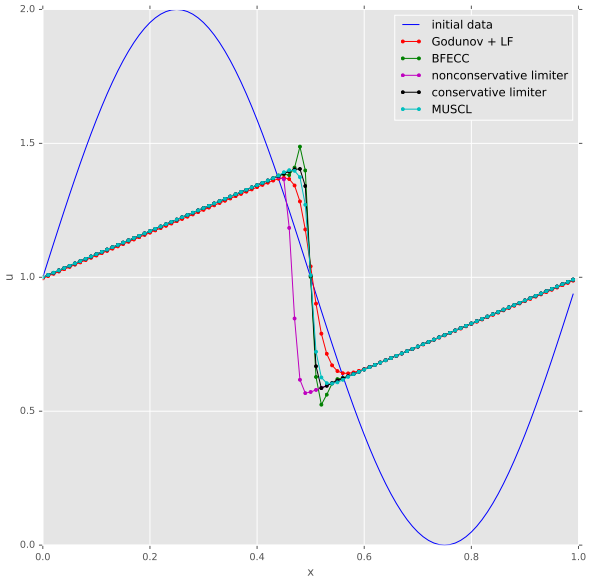


Figure 4.7: Shock tracking for the Burgers equation, comparison 1

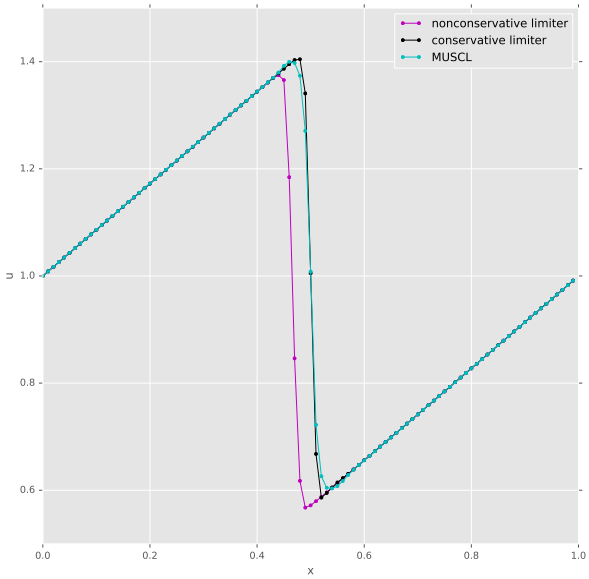


Figure 4.8: Shock tracking for the Burgers equation, comparison 2

4.6 A characteristic difference BFECC scheme for convection-diffusion equation, viscous Burgers equation and KdV equation

Diffusion term shows up in the viscid Burgers' equation

$$u_t + uu_x = \nu u_{xx} \tag{4.7}$$

And the numerical scheme for the viscid Burgers' equation needs to handle the diffusion term properly. In this section, we discuss a characteristic difference BFECC scheme for the viscid Burgers' equation that is second order accurate and unconditionally stable. We also apply the same method for to obtain a second order accurate scheme for the Korteweg-de Vries (KdV) equation.

We first discuss the convection-diffusion equation with constant coefficients

$$u_t + au_x = cu_{xx} \tag{4.8}$$

Note the left hand side is the derivative along the characteristic. Consider a change of variables

$$s = x - at$$
$$\tau = \sqrt{1+a^2}t$$

Define $v(\tau, s) = u(\tau/\sqrt{1+a^2}, s + a\tau/\sqrt{1+a^2})$ (i.e. $u(t,x) = v(\sqrt{1+a^2}t, x-at)$), then v satisfies

$$\frac{\partial v}{\partial \tau} = \tilde{c}\frac{\partial^2 v}{\partial s^2} \tag{4.9}$$

where $\tilde{c} = \frac{c}{\sqrt{1+a^2}}$. Note the left hand side is the derivative along the characteristic.

For a convection-diffusion with variable coefficients, we can no longer convert it into a diffusion equation by a simple transform. Instead we will construct a finite difference scheme based on differencing in the characteristic direction. For equation-4.8 with variable coefficients, we approximate the derivative along the characteristic direction by

$$\frac{U_i^{n+1} - \tilde{U}_i^{n+1}}{\sqrt{1+(a_i^*)^2}\Delta t}$$

where $\tilde{U}_i^{n+1}$ is the value of u at the intersection of time level $t = t_n$ and the characteristic curve going through (t_{n+1}, x_i), and $*$ is the temporal index for a_i (for example, it could be n, $n+1$ or $n+\frac{1}{2}$), we will specify it once we choose an approximation for the spatial second derivative.

Using BFECC + CIR scheme, we can get a second order accurate estimate of $\tilde{u}_i^n$. For

the spatial derivative, we can use central difference approximation at time level t_n (corresponding to an explicit scheme), at t_{n+1} (corresponding to an implicit scheme), or a convex combination of central difference approximation at t_n and t_{n+1} (corresponding to the θ scheme, or the Crank-Nicolson scheme for $\theta = \frac{1}{2}$). The $*$ in the characteristic derivative are then n, $n+1$ or $(1-\theta)n + \theta(n+1)$, respectively.

Using the BFECC + CIR scheme for $\tilde{u}_i^{n+1}$ and central difference at $t_{n+\frac{1}{2}}$ (i.e. Crank-Nicolson type), we get the following scheme for equation-4.8

$$\frac{U_i^{n+1} - \tilde{U}_i^{n+1}}{\sqrt{1+(a_i^{n+1})^2}\Delta t} = \frac{c}{2\sqrt{1+(a_i^{n+1})^2}} \left[\frac{U_{i-1}^{n+1} - 2U_i^{n+1} + U_{i+1}^{n+1}}{\Delta x^2} + \frac{\tilde{U}_{i-1}^{n+1} - 2\tilde{U}_i^{n+1} + \tilde{U}_{i+1}^{n+1}}{\Delta x^2} \right]$$

Cancel $\frac{1}{\sqrt{1+(a_i^{n+1})^2}}$, we get

$$\frac{U_i^{n+1} - \tilde{U}_i^{n+1}}{\Delta t} = \frac{c}{2} \left[\frac{U_{i-1}^{n+1} - 2U_i^{n+1} + U_{i+1}^{n+1}}{\Delta x^2} + \frac{\tilde{U}_{i-1}^{n+1} - 2\tilde{U}_i^{n+1} + \tilde{U}_{i+1}^{n+1}}{\Delta x^2} \right]$$

Therefore, a update step of the characteristic difference BFECC scheme consists of two steps:

1. Solve $u_t + au_x = 0$ from t_n to t_{n+1} with BFECC + CIR scheme, and obtain $\tilde{U}^{n+1}$;

2. Solve tridiagonal linear system, and obtain U^{n+1};

$$\frac{U_i^{n+1} - \tilde{U}_i^{n+1}}{\Delta t} = \frac{c}{2} \left[\frac{U_{i-1}^{n+1} - 2U_i^{n+1} + U_{i+1}^{n+1}}{\Delta x^2} + \frac{\tilde{U}_{i-1}^{n+1} - 2\tilde{U}_i^{n+1} + \tilde{U}_{i+1}^{n+1}}{\Delta x^2} \right]$$

Note both steps are unconditionally stable in l^2 sense, so the scheme is unconditionally stable in l^2 sense. Note although the second step satisfies maximum principle, but the first step usually is not, so it is not guaranteed to satisfies the maximum principle.

For order of accuracy analysis, we first modify one step of the scheme as follows:

1. Solve $u_t + au_x = 0$ from t_n to $t_{n+\frac{1}{2}}$ with BFECC + CIR scheme, and obtain $\tilde{U}^{n+\frac{1}{2}}$;

2. Solve tridiagonal linear system, and obtain $\tilde{U}^{n+1}$;

$$\frac{U_i^{n+1} - \tilde{U}_i^{n+1/2}}{\Delta t} = \frac{c}{2}\left[\frac{U_{i-1}^{n+1} - 2U_i^{n+1} + U_{i+1}^{n+1}}{\Delta x^2} + \frac{\tilde{U}_{i-1}^{n+1/2} - 2\tilde{U}_i^{n+1/2} + \tilde{U}_{i+1}^{n+1/2}}{\Delta x^2}\right]$$

3. Solve $u_t + au_x = 0$ from $t_{n+\frac{1}{2}}$ to t_{n+1} with BFECC + CIR scheme, and obtain U^{n+1}.

This will have the same order of accuracy as the previous step since we can combine the BFECC stages from two adjacent steps into one and only increase the error by at most a factor of 2. Therefore, the order of accuracy doesn't change.

Note this three-stage algorithm is simply the Strang splitting of the convection-diffusion equation-4.8. Since each stage is second order accurate and the Strang splitting is second order accurate, we conclude the characteristic difference BFECC scheme is second order accurate.

A numerical example is solved to verify the effectiveness of the scheme. In this example, we solve convection-diffusion equation $u_t + au_x = cu_{xx}$ on $x \in [0, 1]$ with periodic boundary condition. Here we choose $a = 1.0$ and $c = 0.2$, fix the $\Delta t/\Delta x$ to be 0.5, and set initial condition $u(0, x) = \cos(2\pi x)$. The three stage scheme-[label needed] and the two stage scheme-[label needed] are used to numerically solve the problem on a uniform grid with $20, 40, 80, 160, 320$ grid points, and the numerical solutions are compared to the analytical solution $u(t, x) = e^{-4\pi^2 ct}\cos(2\pi(x - at))$.

The error and order of accuracy are collected in Table-4.3.

Table 4.3: Order of accuracy for convection-diffusion equation, solution at $T = 0.2$.

	Three-stage scheme		Two-stage scheme	
Grid	Error	Order	Error	Order
20	1.491×10^{-3}	–	8.965×10^{-4}	–
40	3.855×10^{-4}	1.95	2.546×10^{-4}	1.82
80	9.836×10^{-5}	1.97	6.758×10^{-5}	1.91
160	2.486×10^{-5}	1.98	1.740×10^{-5}	1.96
320	6.250×10^{-6}	1.99	4.412×10^{-6}	1.98

The characteristic difference BFECC schemes can also be applied to the viscous Burgers' equation $u_t + uu_x = \nu u_{xx}$. Here $\nu \geq 0$ is the viscosity. The three stage BFECC scheme now has a step consisting of the following three stages:

1. Solve $u_t + uu_x = 0$ from t_n to $t_{n+\frac{1}{2}}$ with a BFECC scheme, and obtain $\tilde{U}^{n+\frac{1}{2}}$;

2. Solve tridiagonal linear system, and obtain $\tilde{U}^{n+1}$;

$$\frac{U_i^{n+1} - \tilde{U}_i^{n+1/2}}{\Delta t} = \frac{c}{2}\left[\frac{U_{i-1}^{n+1} - 2U_i^{n+1} + U_{i+1}^{n+1}}{\Delta x^2} + \frac{\tilde{U}_{i-1}^{n+1/2} - 2\tilde{U}_i^{n+1/2} + \tilde{U}_{i+1}^{n+1/2}}{\Delta x^2}\right]$$

3. Solve $u_t + au_x = 0$ from $t_{n+\frac{1}{2}}$ to t_{n+1} with the same BFECC scheme as in stage-1, and obtain U^{n+1}.

Here the BFECC scheme in stage-1 and stage-3 can be a BFECC + MUSCL scheme as discussed in section-4.5. Here we need to interpret U_i^n properly in order to make sense of the three-stage scheme. In stage-1 and stage-3, we interpret U_i^n as the approximated cell average $\frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} u(x, t_n)dx$ in order to use the finite volume MUSCL scheme. In stage-2, we interpret U_i^n as the approximated grid value $u(x_i, t_n)$ in order to use the finite difference Crank-Nicolson scheme. Recall in the MUSCL scheme, a piecewise constant reconstruction is done in each cell, so the cell average interpretation in stage-1/3 and the

grid value interpretation in stage-2 are actually compatible. When $\nu \neq 0$, another choice for the BFECC scheme in stage-1/3 is the BFECC + CIR scheme (with a variable convection velocity field). Note although the BFECC + CIR scheme is not conservative for variable velocity field, it can be applied to the viscous Burgers' equation. Due to the viscosity term νu_{xx}, when $\nu > 0$, there is no discontinuity in the solution of the viscous Burgers' equation. In this case, a scheme don't have to be formally conservative. As long as it is compatible with the equation and has enough accuracy, the conserved quantity would be approximately preserved with similar order of accuracy as the the scheme itself.

Similarly, since solutions to viscous Burgers' equation are smooth when $\nu \neq 0$. Limiting is not necessary. Limiting, however, can improve the stability of the BFECC stage of the scheme and help reduce the numerical error.

We solve the viscous Burgers' equation in the following example using 4 different BFECC schemes for stage-1/3 – BFECC + MUSCL, BFECC + MUSCL + conservative limiter, BFECC + CIR, BFECC + CIR + limiter. Here the viscosity is set to be $\nu = 0.2$, $\Delta t/\Delta x$ is fixed to be 0.2, and initial data is

$$u(0,x) = -2\nu \frac{\frac{1}{2}\omega \cos(\omega x)}{1 + \frac{1}{2}\sin(\omega x)}$$

and exact solution is

$$u(t,x) = -2\nu \frac{\frac{1}{2}\omega e^{-c\omega^2 t} \cos(\omega x)}{1 + \frac{1}{2}e^{-c\omega^2 t}\sin(\omega x)}$$

with $\omega = 4\pi$. Here the exact solution is obtained by applying the Cole-Hopf transformation [39, 40] $u = -2\nu \frac{\partial \ln(\phi)}{\partial x}$ to the Burgers' equation and solving the resultant diffusion equation $\phi_t = \nu \phi_{xx}$ with initial condition $\phi(0,x) = 1 + \frac{1}{2}\sin(\omega x)$.

The grid refinement result is in Table-4.4 and 4.5.

Table 4.4: Order of accuracy for viscous Burgers' equation, solution at $T = 0.2$.

	BFECC + Godunov		BFECC + Godunov + limiter	
Grid	Error	Order	Error	Order
20	5.098×10^{-3}	–	3.091×10^{-3}	–
40	6.859×10^{-4}	2.89	1.592×10^{-3}	0.95
80	9.099×10^{-4}	-0.41	9.113×10^{-4}	0.81
160	5.774×10^{-4}	0.65	2.527×10^{-4}	1.85
320	3.172×10^{-4}	0.86	9.006×10^{-5}	1.49

Table 4.5: Order of accuracy for viscous Burgers' equation, solution at $T = 0.2$.

	BFECC + CIR		BFECC + CIR + limiter	
Grid	Error	Order	Error	Order
20	5.700×10^{-4}	–	5.700×10^{-4}	–
40	1.371×10^{-4}	2.06	1.371×10^{-4}	2.06
80	3.407×10^{-5}	2.01	3.407×10^{-5}	2.01
160	8.525×10^{-6}	2.00	8.525×10^{-6}	2.00
320	2.134×10^{-6}	2.00	2.134×10^{-6}	2.00

The same method can be applied to the Korteweg-de Vries (KdV) equations. Consider the KdV equation

$$u_t + 6uu_x + \nu u_{xxx} = 0 \tag{4.10}$$

Here ν is usually set to be 1. We can apply the characteristic difference idea to this equation, solve the advection part $u_t + 6uu_x = 0$ with a BFECC scheme (for example, BFECC + CIR or BFECC + Godunov) and solve the dispersion part $u_t + \nu u_{xxx} = 0$ with a Crank-Nicolson type scheme. In the following, we use a scheme proposed by Kruskal [41] to

solve $u_t + \nu u_{xxx} = 0$

$$\frac{U_i^{n+1} - U_i^n}{\Delta t} + \nu \frac{U_{i+2}^{n+1} - 3U_{i+1}^{n+1} + 3U_i^{n+1} - U_{i-1}^{n+1}}{2(\Delta x)^3} + \nu \frac{U_{i+1}^n - 3U_i^n + 3U_{i-1}^n - U_{i-2}^n}{2(\Delta x)^3} = 0 \tag{4.11}$$

In the following example, we solve the KdV equation-4.10 on $x \in [0, 1]$ with periodic boundary condition. Here we set $\nu = 1$ and the initial data to be a soliton solution at $t = 0$:

$$u(0, x) = \frac{c}{2} \operatorname{sech}^2 \left[\frac{\sqrt{c}}{2} (x - 0.4) \right]. \tag{4.12}$$

Note the exact soliton solution is

$$u(t, x) = \frac{c}{2} \operatorname{sech}^2 \left[\frac{\sqrt{c}}{2} (x - ct - 0.4) \right]. \tag{4.13}$$

Here we set $c = 2000$ in order to make the soliton solution supported in $[0, 1]$. Numerical solutions at $T = 10^{-4}$ and exact solution-(4.13) are plotted in Figure-4.6. $T = 10^{-4}$ is selected so that the soliton is still confined in the $[0, 1]$ (note the soliton propagate at speed c). The grid refinement analysis is shown in Table-4.6 and 4.7. We see the method achieves second order accuracy.

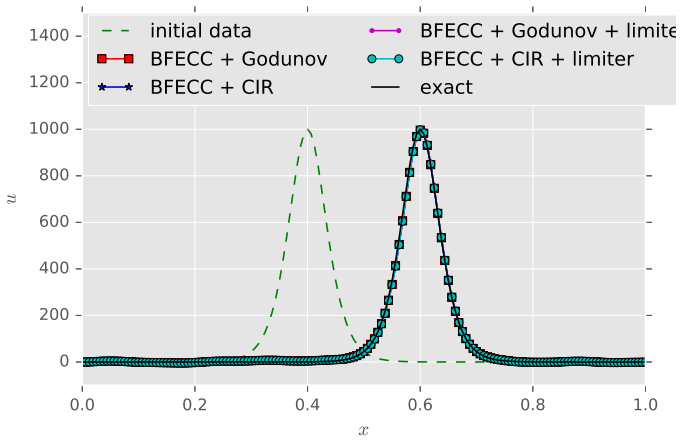


Figure 4.9: Propagation of a soliton solution for KdV equation.

Table 4.6: l^2 error and order of accuracy. KdV equation, solution at $T = 10^{-4}$.

	BFECC + Godunov		BFECC + Godunov + limiter	
Grid	Error	Order	Error	Order
20	1.754×10^2	–	1.721×10^2	–
40	1.134×10^2	0.63	1.075×10^2	0.67
80	3.406×10^{-1}	1.74	2.837×10^1	1.92
160	6.285	2.44	6.568	2.11
320	2.098	1.58	2.122	1.63

Table 4.7: l^2 error and order of accuracy. KdV equation, solution at $T = 10^{-4}$.

	BFECC + CIR		BFECC + CIR + limiter	
Grid	Error	Order	Error	Order
20	2.865×10^2	–	2.546×10^2	–
40	2.882×10^2	-0.008	2.521×10^2	0.014
80	7.763×10^1	1.89	7.646×10^1	1.72
160	5.835	3.73	5.792	3.72
320	1.373	2.08	1.375	2.08

The following numerical example shows the long time performance of the scheme. We solve a modified KdV equation $u_t + uu_x + \nu u_{xxx} = 0$ with $\nu = 4.84 \times 10^{-4}$. Numerical solutions are calculated with BFECC + CIR + limiter as the solver for the convection term. Here the initial condition is a cosine profile $u(0, x) = \cos(2\pi x)$, boundary condition is the periodic boundary condition and $\Delta t/\Delta x = 0.25$. Numerical solution at $t = 0, 0.125, 0.25, 0.375, 0.5, 0.625$ are shown in Figure 4.6. We observe the sinusoidal initial data evolves into solitons with different magnitudes and propagation speed. This agrees with the asymptotic behavior of solutions for the KdV equations, where smooth solutions evolves into solitons when $t \to \infty$ [42, 43].

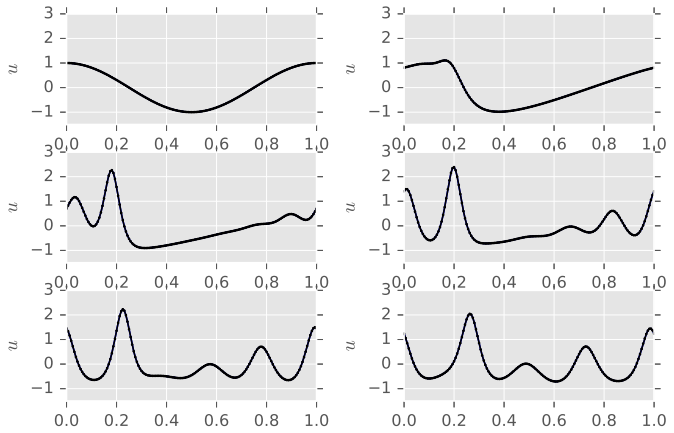


Figure 4.10: Cosine solution evolves into solitons: from upper left to lower right, the figures shows numerical solutions at $t = 0, 0.125, 0.25, 0.375, 0.5$ and 0.625.

To sum up this chapter, we note all the equations in this chapter arises from conservation laws. The numerical examples demonstrate the effectiveness of BFECC method and the conservative limiter for scalar conservation laws. When viscosity or dispersion is presented, the BFECC + CIR + limiter serves as an effective solver for the convection part of the equation.

CHAPTER 5
CONCLUSION

We studied two extensions of the BFECC method in this thesis: BFECC method for linear hyperbolic PDE systems, and BFECC method for nonlinear scalar conservation laws.

The BFECC method for linear hyperbolic PDE systems is a natural extension to BFECC method for scalar advection equations. We showed the same stability improvement is valid under modest assumptions for the systems and underlying scheme. In particular, central difference scheme can be made stable when the BFECC method is applied, and Lax-Friedrichs schemes combined with BFECC method has a larger CFL number. For symmetric linear hyperbolic systems with constant coefficients, we proved the order improvement for odd ordered underlying scheme. This is particularly useful for the Maxwell's equations.

Based on the BFECC method for linear hyperbolic systems, we proposed BFECC schemes for the Maxwell's equations. The schemes use first order accurate central difference and Lax-Friedrichs schems as underlying scheme. The BFECC method boosts the order of accuracy to second order, and increases the CFL numbers to $\sqrt{3}$ and 2, respectively. On non-orthogonal unstructured grids, we propose to use a simple least square approximation scheme as the underlying scheme. The BFECC method combined with least square approximation schemes yields new schemes that are second order accurate on non-orthogonal unstructured grids. Numerical experiments are carried out to study the performance of the schemes on uniform rectangular grids, non-rectangular grids, and for scattering problems that have variable coefficients. All the experiments confirm the second order accuracy and numerical stabilities of the proposed schemes. Compared to FDTD schemes on non-orthogonal or unstructured grids such as Nonorthogonal FDTD scheme or Generalized Yee scheme, the BFECC schemes don't require generation of staggered non-orthogonal grids, is stable for long time simulation and simple to implement. On uniform

orthogonal grids, it has larger CFL number than the Yee scheme, but is less efficient.

For nonlinear scalar conservation laws, we studied a limiter that reduces spurious oscillations near discontinuities and guarantees to be conservative when the underlying scheme is conservative. The BFECC method is extended to finite volume schemes so that classical schemes such as the Godunov scheme can be used as the underlying scheme. BFECC + Godunov scheme + conservative limiter shows better numerical results than the classical MUSCL scheme, demonstrating the potential of BFECC methods for conservation laws. The BFECC method is also successfully applied to convection terms in Vlasov-Poisson, viscous Burgers equation and the KdV equation.

The success in applying BFECC method to linear hyperbolic systems and nonlinear scalar conservation laws motivates more study of BFECC method for nonlinear systems of conservation laws, such as the Euler equations. Some interesting directions that are worth pursuing are:

1. Applying the BFECC schemes for the Maxwell's equations to more problems. We studied the BFECC schemes on uniform orthogonal grids, for which we provided theorems to guarantee stability and second order accuracy. We also tested the BFECC schemes on non-orthogonal grids and demonstrated stability and accuracy on smoothly varying non-orthogonal grids or point shifted grids. It is interesting to see more studies on their performance on different types of grids, and on special geometric structures such as wires [6]. Our numerical examples on scattering problems suggest the BFECC schemes work well for the Maxwell's equations with variable coefficients. More studies on problems with variable coefficients and material interfaces are needed before the BFECC scheme can compete with classical local subcell models [6].

2. Applying the conservative limiters to nonlinear systems of conservation laws, such as the Euler equations. The current conservative limiters have been tested for scalar

conservation laws for more than one spatial dimension with good results (not reported in this thesis). For systems of conservation laws, the limiter may need to be modified, since oscillation in one variable may be caused by other variables. Proper conservative limiter for system of conservation laws are an interesting topic to pursue in future.

3. Rigorous proof for accuracy improvement on non-orthogonal grids. Our current analysis for the stability and accuracy improvement of the BFECC method is based on discrete Fourier transform, and it not applicable to non-orthogonal grids. The numerical examples suggests BFECC method also improve stability and accuracy for schemes on non-orthogonal grids. Providing rigorous analysis for the stability and accuracy of BFECC schemes would be a challenging topic for future work.

Appendices

APPENDIX A
STABILITY AND ACCURACY OF BFECC SCHEMES BASED ON CENTRAL DIFFERENCE

In this appendix, we discuss in detail the calculation for the Fourier symbol matrices of the central difference scheme and BFECC based on the central difference schemes in one and two dimensional case. The Fourier symbol matrices confirm that BFECC based on the central difference scheme is second order accurate.

A.1 One dimensional case

For Maxwell's equations in one dimensional free space with periodic boundary condition, central difference scheme $\mathcal{L}$'s Fourier symbol matrix is

$$Q_{\mathcal{L}} = \begin{pmatrix} 1 & i\lambda \sin(2\pi kh) \\ i\lambda \sin(2\pi kh) & 1 \end{pmatrix},$$

and $\mathcal{L}^*$'s Fourier symbol matrix is $Q_{\mathcal{L}^*} = \overline{Q_{\mathcal{L}}}$.

For BFECC based on the central difference scheme, its Fourier symbol matrix is

$$Q_B = Q_{\mathcal{L}} \left(I + \frac{1}{2}(I - Q_{\mathcal{L}^*} Q_{\mathcal{L}}) \right) = \left(1 - \frac{1}{2}\lambda^2 \sin^2(2\pi kh) \right) \begin{pmatrix} 1 & i\lambda \sin(2\pi kh) \\ i\lambda \sin(2\pi kh) & 1 \end{pmatrix}$$

Stability We calculate eigenvalues for $Q_{\mathcal{L}}$ and Q_B

$$\lambda(Q_{\mathcal{L}})_{\pm} = 1 \pm i\lambda \sin(2\pi kh)$$
$$\lambda(Q_B)_{\pm} = \left(1 - \frac{1}{2}\lambda^2 \sin^2(2\pi kh) \right) (1 \pm i\lambda \sin(2\pi kh))$$

We study the spectral radius of Q_B:

$$|\lambda(Q_B)_\pm|^2 = \left(1 - \frac{1}{2}\lambda^2 \sin^2(2\pi kh)\right)^2 (1 + \lambda^2 \sin^2(2\pi kh))$$

Let $\zeta = \sin^2(2\pi kh) \in [0, 1]$, and define

$$f(\zeta) = |\lambda(Q_B)_\pm|^2 = \left(1 - \frac{1}{2}\lambda^2 \zeta\right)^2 (1 + \lambda^2 \zeta)$$

When $\lambda^2 \leq 2$, $f(\zeta)$ is monotonically decreasing in $[0, 1]$, and it obtains its maximum at 0, $f(0) = 1$ and for all $\zeta \in (0, 1]$, $f(\zeta) < 1$. For the case $f(0) = 1$, we can explicitly check that mode is stable. Therefore for $\lambda^2 \leq 2$, the scheme is stable.

When $\lambda^2 > 2$, $\max_{\zeta\in[0,1]} f(\zeta) = \max(f(0), f(1)) = \max\left(1, \left(1 - \frac{1}{2}\lambda^2\right)^2 (1 + \lambda^2)\right)$, we already checked $k = 0$ is always a stable mode. Setting $\left(1 - \frac{1}{2}\lambda^2\right)^2 (1 + \lambda^2) < 1$, we get $\lambda^2 < 3$.

Therefore $\Delta t/\Delta x = \lambda < \sqrt{3}$ ensures l^2 stability for the BFECC scheme.

Accuracy Write

$$E(t, x) = \sum_{k\in F_N} C_k(t) e^{2\pi ikx}$$
$$H(t, x) = \sum_{k\in F_N} D_k(t) e^{2\pi ikx}$$

and plug in the Maxwell's equations, we get:

$$\frac{d}{dt}\begin{pmatrix} C_k \\ D_k \end{pmatrix} = \begin{pmatrix} 0 & 2\pi ik \\ 2\pi ik & 0 \end{pmatrix}\begin{pmatrix} C_k \\ D_k \end{pmatrix} = G\begin{pmatrix} C_k \\ D_k \end{pmatrix}$$

where matrix G is defined by the last equality. Calculate the matrix exponential, we get:

$$\begin{pmatrix} C_k(t_n+\Delta t) \\ D_k(t_n+\Delta t) \end{pmatrix} = e^{\Delta t G} \begin{pmatrix} C_k(t_n) \\ D_k(t_n) \end{pmatrix} = \begin{pmatrix} \cos(2\pi k\Delta t) & i\sin(2\pi k\Delta t) \\ i\sin(2\pi k\Delta t) & \cos(2\pi k\Delta t) \end{pmatrix} \begin{pmatrix} C_k(t_n) \\ D_k(t_n) \end{pmatrix}$$

While with BFECC + central difference, we have:

$$\begin{pmatrix} C_k(t_n+\Delta t) \\ D_k(t_n+\Delta t) \end{pmatrix} = Q_B \begin{pmatrix} C_k(t_n) \\ D_k(t_n) \end{pmatrix} = \left(1-\frac{1}{2}\lambda^2\sin^2(2\pi kh)\right) \begin{pmatrix} 1 & i\lambda\sin(2\pi kh) \\ i\lambda\sin(2\pi kh) & 1 \end{pmatrix} \begin{pmatrix} C_k(t_n) \\ D_k(t_n) \end{pmatrix}$$

Note $\lambda h = \Delta t$, we see:

$$Q_B = e^{\Delta t G} + O(|kh|^3), \text{ as } h \to 0$$

By Theorem-2 in Chapter 2, we see the BFECC + central difference scheme is a second order accurate scheme.

A.2 Two dimensional case

For Maxwell's equations in two dimensional free space with periodic boundary condition, central difference scheme $\mathcal{L}$'s Fourier symbol matrix is

$$Q_{\mathcal{L}} = \begin{pmatrix} 1 & 0 & -i\lambda_y\sin(2\pi l\Delta y) \\ 0 & 1 & i\lambda_x\sin(2\pi k\Delta x) \\ -i\lambda_y\sin(2\pi l\Delta y) & i\lambda_x\sin(2\pi k\Delta x) & 1 \end{pmatrix}$$

As discussed in section 3.2, $Q_{\mathcal{L}^*} = \overline{Q_{\mathcal{L}}}$. For convenience of notation, we denote $s_k^x = \sin(2\pi k\Delta x)$ and $s_l^y = \sin(2\pi l\Delta y)$.

The BFECC + Central Difference scheme has Fourier symbol matrix

$$Q_B = Q_{\mathcal{L}}\left(I + \frac{1}{2}(I - Q_{\mathcal{L}^*}Q_{\mathcal{L}})\right)$$

By direct computation, we get

$$I+\frac{1}{2}(I-Q_{\mathcal{L}^*}Q_{\mathcal{L}})=\begin{pmatrix} 1-\frac{1}{2}\lambda_y^2(s_l^y)^2 & \frac{1}{2}\lambda_x\lambda_y s_k^x s_l^y & 0 \\ \frac{1}{2}\lambda_x\lambda_y s_k^x s_l^y & 1-\frac{1}{2}\lambda_x^2(s_k^x)^2 & 0 \\ 0 & 0 & 1-\frac{1}{2}\lambda_x^2(s_k^x)^2-\frac{1}{2}\lambda_y^2(s_l^y)^2 \end{pmatrix}$$

and

$$Q_B=\begin{pmatrix} 1-\frac{1}{2}\lambda_y^2(s_l^y)^2 & \frac{1}{2}\lambda_x\lambda_y s_k^x s_l^y & -i\lambda_y s_l^y\left(1-\frac{1}{2}\lambda_x^2(s_k^x)^2-\frac{1}{2}\lambda_y^2(s_l^y)^2\right) \\ \frac{1}{2}\lambda_x\lambda_y s_k^x s_l^y & 1-\frac{1}{2}\lambda_x^2(s_k^x)^2 & i\lambda_x s_k^x\left(1-\frac{1}{2}\lambda_x^2(s_k^x)^2-\frac{1}{2}\lambda_y^2(s_l^y)^2\right) \\ -i\lambda_y s_l^y\left(1-\frac{1}{2}\lambda_x^2(s_k^x)^2-\frac{1}{2}\lambda_y^2(s_l^y)^2\right) & i\lambda_x s_k^x\left(1-\frac{1}{2}\lambda_x^2(s_k^x)^2-\frac{1}{2}\lambda_y^2(s_l^y)^2\right) & 1-\frac{1}{2}\lambda_x^2(s_k^x)^2-\frac{1}{2}\lambda_y^2(s_l^y)^2 \end{pmatrix} \quad \text{(A.1)}$$

Stability Eigenvalues of $Q_{\mathcal{L}}$ are

$$\lambda_1=1,\ \lambda_{2,3}=1\pm i\sqrt{\lambda_x^2(s_k^x)^2+\lambda_y^2(s_l^y)^2}$$

The matrix A can be decomposed as

$$Q_{\mathcal{L}}=V\Lambda V^{-1}$$

where

$$\Lambda=\text{diag}\{\lambda_1,\lambda_2,\lambda_3\}$$

and

$$V=\begin{pmatrix} \lambda_x s_k^x & -\lambda_y s_l^y & \lambda_y s_l^y \\ \lambda_y s_l^y & \lambda_x s_k^x & -\lambda_x s_k^x \\ 0 & \sqrt{\lambda_x^2(s_k^x)^2+\lambda_y^2(s_l^y)^2} & \sqrt{\lambda_x^2(s_k^x)^2+\lambda_y^2(s_l^y)^2} \end{pmatrix}$$

It is then easy to see the scheme is l^2 stable if and only if $\max_{s_k^x,s_l^y}|\lambda_{2,3}|\leq 1$, which is

not true since $|\lambda_{2,3}|^2 = 1 + \lambda_x^2(s_k^x)^2 + \lambda_y(s_l^y)^2 > 1$ for $s_k^x \neq 0$ or $s_l^y \neq 0$. Therefore the central difference scheme is unconditionally unstable.

We can verify that columns of V are also eigenvectors of $\overline{Q_{\mathcal{L}}}Q_{\mathcal{L}}$ and hence eigenvectors of Q_B. This allows us to compute the eigenvalues of Q_B:

$$\lambda_1(Q_B) = 1, \lambda_{2,3}(Q_B) = \left(1 - \frac{1}{2}(\lambda_x^2(s_k^x)^2 + \lambda_y^2(s_l^y)^2)\right)\left(1 \pm i\sqrt{\lambda_x^2(s_k^x)^2 + \lambda_y^2(s_l^y)^2}\right)$$

Therefore, the BFECC + Central Difference scheme is stable if and only if

$$\max_{s_k^x, s_l^y}\left(1 - \frac{1}{2}(\lambda_x^2(s_k^x)^2 + \lambda_y^2(s_l^y)^2)\right)^2\left(1 + \lambda_x^2(s_k^x)^2 + \lambda_y^2(s_l^y)^2\right) \leq 1$$

Let $\zeta = (s_k^x)^2 \in [0,1], \theta = (s_l^y)^2 \in [0,1]$, define

$$f(\zeta, \theta) = \left(1 - \frac{1}{2}(\lambda_x^2\zeta + \lambda_y^2\theta)\right)^2\left(1 + \lambda_x^2\zeta + \lambda_y^2\theta\right)$$

We have

$$\frac{\partial f}{\partial \zeta} < 0$$
$$\frac{\partial f}{\partial \theta} < 0$$

when $1 - \frac{1}{2}(\lambda_x^2\zeta + \lambda_y^2\theta) > 0$ and above this line, both partial derivatives are positive.

Using this property, we see

$$\max_{0\leq\zeta,\theta\leq 1} f(\zeta, \theta) = f(0,0) = 1, \text{ if } \lambda_x^2 + \lambda_y^2 < 2$$

$$\max_{0\leq\zeta,\theta\leq 1} f(\zeta, \theta) = \max(f(0,0), f(1,1)), \text{ if } \lambda_x^2 + \lambda_y^2 > 2$$

For the case, $\lambda_x^2 + \lambda_y^2 > 2$ the l^2 stability condition becomes

$$f(1,1) = \left(1 - \frac{1}{2}(\lambda_x^2 + \lambda_y^2)\right)^2 \left(1 + \lambda_x^2 + \lambda_y^2\right) \leq 1 \Leftrightarrow \lambda_x^2 + \lambda_y^2 \leq 3$$

Therefore, the BFECC + Central Difference scheme is stable is stable if and only if

$$\lambda_x^2 + \lambda_y^2 \leq 3 \Leftrightarrow \Delta t \leq \frac{\sqrt{3}}{\sqrt{(1/\Delta x)^2 + (1/\Delta y)^2}}$$

Accuracy Write

$$\begin{aligned} H_x &= \sum_{k,l \in \mathcal{F}_N} C_{k,l}(t) e^{2\pi i(kx+ly)} \\ H_y &= \sum_{k,l \in \mathcal{F}_N} D_{k,l}(t) e^{2\pi i(kx+ly)} \\ E_z &= \sum_{k,l \in \mathcal{F}_N} E_{k,l}(t) e^{2\pi i(kx+ly)} \end{aligned}$$

Plug into the Maxwell's equations and get

$$\frac{\partial}{\partial t} \begin{pmatrix} C_{k,l} \\ D_{k,l} \\ E_{k,l} \end{pmatrix} = \begin{pmatrix} 0 & 0 & -2\pi il \\ 0 & 0 & 2\pi ik \\ -2\pi il & 2\pi ik & 0 \end{pmatrix} \begin{pmatrix} C_{k,l} \\ D_{k,l} \\ E_{k,l} \end{pmatrix} = G \begin{pmatrix} C_{k,l} \\ D_{k,l} \\ E_{k,l} \end{pmatrix}$$

Calculate the matrix exponential to get

$$\begin{pmatrix} C_{k,l}(t+\Delta t) \\ D_{k,l}(t+\Delta t) \\ E_{k,l}(t+\Delta t) \end{pmatrix} = e^{\Delta t G} \begin{pmatrix} C_{k,l}(t) \\ D_{k,l}(t) \\ E_{k,l}(t) \end{pmatrix}$$

where

$$e^{\Delta tG} = \begin{pmatrix} \frac{k^2+l^2\cos(2\pi\sqrt{k^2+l^2}\Delta t)}{k^2+l^2} & \frac{kl(1-\cos(2\pi\sqrt{k^2+l^2}\Delta t))}{k^2+l^2} & -i\frac{l\sin(2\pi\sqrt{k^2+l^2}\Delta t)}{\sqrt{k^2+l^2}} \\ \frac{kl(1-\cos(2\pi\sqrt{k^2+l^2}\Delta t))}{k^2+l^2} & \frac{l^2+k^2\cos(2\pi\sqrt{k^2+l^2}\Delta t)}{k^2+l^2} & i\frac{k\sin(2\pi\sqrt{k^2+l^2}\Delta t)}{\sqrt{k^2+l^2}} \\ -i\frac{l\sin(2\pi\sqrt{k^2+l^2}\Delta t)}{\sqrt{k^2+l^2}} & i\frac{k\sin(2\pi\sqrt{k^2+l^2}\Delta t)}{\sqrt{k^2+l^2}} & \cos(2\pi\sqrt{k^2+l^2}\Delta t) \end{pmatrix}$$

Note it is symmetric. Expand entries of $e^{\Delta tG}$ upto second order, the entries are listed as (in the order of $(1,1),(1,2),(1,3),(2,2),(2,3),(3,3)$):

$$\begin{aligned}
&\frac{k^2+l^2\cos(2\pi\sqrt{k^2+l^2}\Delta t)}{k^2+l^2} = 1-\frac{1}{2}(2\pi l^2)(\Delta t)^2+O(\Delta t^3)\\
&\frac{kl(1-\cos(2\pi\sqrt{k^2+l^2}\Delta t))}{k^2+l^2} = \frac{1}{2}(2\pi)^2 kl(\Delta t)^2+O(\Delta t^3)\\
&-i\frac{l\sin(2\pi\sqrt{k^2+l^2}\Delta t)}{\sqrt{k^2+l^2}} = -i2\pi l\Delta t+O(\Delta t^3)\\
&\frac{l^2+k^2\cos(2\pi\sqrt{k^2+l^2}\Delta t)}{k^2+l^2} = 1-\frac{1}{2}(2\pi k^2)(\Delta t)^2+O(\Delta t^3)\\
&i\frac{k\sin(2\pi\sqrt{k^2+l^2}\Delta t)}{\sqrt{k^2+l^2}} = i2\pi k\Delta t+O(\Delta t^3)\\
&\cos(2\pi\sqrt{k^2+l^2}\Delta t) = 1-\frac{1}{2}(2\pi)^2(k^2+l^2)(\Delta t)^2+O(\Delta t^3)
\end{aligned}$$

Compare with entries of $Q_{\mathcal{L}}$, we see $Q_{\mathcal{L}} = e^{\Delta tG}+O(|\sqrt{k^2+l^2}\Delta t|^2)$, by Theorem-2 in Chapter 2, the central difference scheme is first order accurate.

Expand entries of Q_B and note $\lambda_x\Delta x = \Delta t$ and $\lambda_y\Delta y = \Delta t$, the entries are listed as

(in the order of $(1,1),(1,2),(1,3),(2,2),(2,3),(3,3)$):

$$
\begin{aligned}
1-\frac{1}{2}\lambda_y^2(s_l^y)^2 &= 1-\frac{1}{2}(2\pi l^2)(\Delta t)^2+O(\Delta t^3)\\
\frac{1}{2}\lambda_x\lambda_y s_k^x s_l^y &= \frac{1}{2}(2\pi)^2 kl(\Delta t)^2+O(\Delta t^3)\\
-i\lambda_y s_l^y\left(1-\frac{1}{2}(\lambda_x^2(s_k^x)^2+\lambda_y^2(s_l^y)^2)\right) &= -i2\pi l\Delta t+O(\Delta t^3)\\
1-\frac{1}{2}\lambda_x^2(s_k^x)^2 &= 1-\frac{1}{2}(2\pi k^2)(\Delta t)^2+O(\Delta t^3)\\
i\lambda_x s_k^x\left(1-\frac{1}{2}(\lambda_x^2(s_k^x)^2+\lambda_y^2(s_l^y)^2)\right) &= i2\pi k\Delta t+O(\Delta t^3)\\
1-\frac{1}{2}(\lambda_x^2(s_k^x)^2+\lambda_y^2(s_l^y)^2) &= 1-\frac{1}{2}(2\pi)^2(k^2+l^2)(\Delta t)^2+O(\Delta t^3)
\end{aligned}
$$

Compare with entries of $e^{\Delta tG}$, we see $Q_B = e^{\Delta tG}+O(|\sqrt{k^2+l^2}\Delta t|^3)$, by Theorem-2 in Chapter 2, the BFECC + central difference scheme is second order accurate.

REFERENCES

[1] F. John, *Partial differential equations, volume 1 of applied mathematical sciences*, 1982.

[2] J. W. Thomas, *Numerical partial differential equations: finite difference methods*. Springer Science & Business Media, 2013, vol. 22.

[3] K. Yee, "Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media," *IEEE Transactions on antennas and propagation*, vol. 14, no. 3, pp. 302–307, 1966.

[4] S. D. Gedney and F Lansing, "Full wave analysis of printed microstrip devices using a generalized yee-algorithm," in *Antennas and Propagation Society International Symposium, 1993. AP-S. Digest*, IEEE, 1993, pp. 1179–1182.

[5] S Gedney, F Lansing, and D Rascoe, "A generalized yee-algorithm for the analysis of mmic devices," *IEEE Transactions on Microwave Theory and Techniques*,

[6] A. Taflove and S. C. Hagness, *Computational electrodynamics: the finite-difference time-domain method*. Artech house, 2005.

[7] S. K. Godunov, "A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics," *Matematicheskii Sbornik*, vol. 89, no. 3, pp. 271–306, 1959.

[8] B. Van Leer, "Towards the ultimate conservative difference scheme. v. a second-order sequel to godunov's method," *Journal of computational Physics*, vol. 32, no. 1, pp. 101–136, 1979.

[9] C.-W. Shu, "Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws," in *Advanced numerical approximation of nonlinear hyperbolic equations*, Springer, 1998, pp. 325–432.

[10] ——, "High order weighted essentially nonoscillatory schemes for convection dominated problems," *SIAM review*, vol. 51, no. 1, pp. 82–126, 2009.

[11] W. H. Reed and T. Hill, "Triangular mesh methods for the neutron transport equation," Los Alamos Scientific Lab., N. Mex.(USA), Tech. Rep., 1973.

[12] S. Bertoluzza, G. Russo, S. Falletta, and C.-W. Shu, "Discontinuous galerkin method for conservation laws," *Numerical Solutions of Partial Differential Equations*, pp. 157–174, 2009.

[13] T. F. Dupont and Y. Liu, "Back and forth error compensation and correction methods for removing errors induced by uneven gradients of the level set function," *Journal of Computational Physics*, vol. 190, no. 1, pp. 311–324, 2003.

[14] ——, "Back and forth error compensation and correction methods for semi-lagrangian schemes with application to level set interface computations," *Mathematics of Computation*, pp. 647–668, 2007.

[15] B. Kim, Y. Liu, I. Llamas, and J. Rossignac, "Flowfixer: using bfecc for fluid simulation," in *Proceedings of the First Eurographics conference on Natural Phenomena*, Eurographics Association, 2005, pp. 51–56.

[16] ——, "Advections with significantly reduced dissipation and diffusion," *IEEE transactions on visualization and computer graphics*, vol. 13, no. 1, 2007.

[17] B. Kim, Y. Liu, I. Llamas, X. Jiao, and J. Rossignac, "Simulation of bubbles in foam with the volume control method," *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3, p. 98, 2007.

[18] A. Selle, R. Fedkiw, B. Kim, Y. Liu, and J. Rossignac, "An unconditionally stable maccormack method," *Journal of Scientific Computing*, vol. 35, no. 2-3, pp. 350–371, 2008.

[19] J.-P. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves," *Journal of computational physics*, vol. 114, no. 2, pp. 185–200, 1994.

[20] M. Zhang and C.-W. Shu, "An analysis of three different formulations of the discontinuous galerkin method for diffusion equations," *Mathematical Models and Methods in Applied Sciences*, vol. 13, no. 03, pp. 395–413, 2003.

[21] ——, "An analysis of and a comparison between the discontinuous galerkin and the spectral finite volume methods," *Computers & fluids*, vol. 34, no. 4-5, pp. 581–592, 2005.

[22] P. D. Lax, "On the stability of difference approximations to solutions of hyperbolic equations with variable coefficients," *Communications on Pure and Applied Mathematics*, vol. 14, no. 3, pp. 497–520, 1961.

[23] P. Monk and E. Süli, "A convergence analysis of yees scheme on nonuniform grids," *SIAM Journal on Numerical Analysis*, vol. 31, no. 2, pp. 393–412, 1994.

[24] P. Monk and E. Suli, "Error estimates for yee's method on non-uniform grids," *IEEE Transactions on Magnetics*, vol. 30, no. 5, pp. 3200–3203, 1994.

[25] R Palandech, R Mittra, *et al.*, "Modeling three-dimensional discontinuities in waveguides using nonorthogonal fdtd algorithm," *IEEE Transactions on Microwave Theory and Techniques*, vol. 40, no. 2, pp. 346–352, 1992.

[26] J. Liu, M. Brio, and J. V. Moloney, "Overlapping yee fdtd method on nonorthogonal grids," *Journal of Scientific Computing*, vol. 39, no. 1, pp. 129–143, 2009.

[27] A. C. Cangellaris and D. B. Wright, "Analysis of the numerical error caused by the stair-stepped approximation of a conducting boundary in fdtd simulations of electromagnetic phenomena," *IEEE transactions on antennas and propagation*, vol. 39, no. 10, pp. 1518–1525, 1991.

[28] T. G. Jurgens, A. Taflove, K. Umashankar, and T. G. Moore, "Finite-difference time-domain modeling of curved surfaces (em scattering)," *IEEE Transactions on Antennas and Propagation*, vol. 40, no. 4, pp. 357–366, 1992.

[29] O. A. McBryan, *Elliptic and hyperbolic interface refinement in two phase flow, in Boundary and Interior Layers*. J. J. H. Miller, ed., Boole Press, Dublin, 1980.

[30] B. Engquist and A. Majda, "Absorbing boundary conditions for numerical simulation of waves," *Proceedings of the National Academy of Sciences*, vol. 74, no. 5, pp. 1765–1766, 1977.

[31] D. Komatitsch and R. Martin, "An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation," *Geophysics*, vol. 72, no. 5, SM155–SM167, 2007.

[32] F. Nataf, *Absorbing boundary conditions and perfectly matched layers in wave propagation problems*, 2013.

[33] J. B. Schneider, "Understanding the finite-difference time-domain method," *School of electrical engineering and computer science Washington State University.–URL: http://www. Eecs. Wsu. Edu/˜ schneidj/ufdtd/(request data: 29.11. 2012)*, 2010.

[34] C. F. Bohren and D. R. Huffman, *Absorption and scattering of light by small particles*. John Wiley & Sons, 2008.

[35] L. Hu, Y. Li, and Y. Liu, "A limiting strategy for the back and forth error compensation and correction method for solving advection equations," *Mathematics of Computation*, vol. 85, no. 299, pp. 1263–1280, 2016.

[36] T. Zhou, Y. Guo, and C.-W. Shu, "Numerical study on landau damping," *Physica D: Nonlinear Phenomena*, vol. 157, no. 4, pp. 322–333, 2001.

[37] F. Filbet and E. Sonnendrücker, "Comparison of eulerian vlasov solvers," *Computer Physics Communications*, vol. 150, no. 3, pp. 247–266, 2003.

[38] J.-M. Qiu and A. Christlieb, "A conservative high order semi-lagrangian weno method for the vlasov equation," *Journal of Computational Physics*, vol. 229, no. 4, pp. 1130–1149, 2010.

[39] J. D. Cole, "On a quasi-linear parabolic equation occurring in aerodynamics," *Quarterly of applied mathematics*, vol. 9, no. 3, pp. 225–236, 1951.

[40] E. Hopf, "The partial differential equation ut+ u ux= mu uxx," *Communications on Pure and Applied mathematics*, vol. 3, no. 3, pp. 201–230, 1950.

[41] T. R. Taha and M. I. Ablowitz, "Analytical and numerical aspects of certain nonlinear evolution equations. iii. numerical, korteweg-de vries equation," *Journal of Computational Physics*, vol. 55, no. 2, pp. 231–253, 1984.

[42] K. Grunert and G. Teschl, "Long-time asymptotics for the korteweg–de vries equation via nonlinear steepest descent," *Mathematical Physics, Analysis and Geometry*, vol. 12, no. 3, pp. 287–324, 2009.

[43] N. J. Zabusky and M. D. Kruskal, "Interaction of" solitons" in a collisionless plasma and the recurrence of initial states," *Physical review letters*, vol. 15, no. 6, p. 240, 1965.